The Sequencing, Assembly and Annotation of Sugar Producing Green Algae and the Design of a Low Cost Turbidostat

A THESIS SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA BY

Matthew B. Arriola

IN PARTIAL FULLFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

Advisor: Dr. Brett M. Barney

October 2017

Acknowledgements

I would like to thank my advisor Prof. Barney for giving me the opportunity to work in his lab, the support necessary to accomplish this work and the freedom to let me try new things. I would like to thank Prof. Ishii and Prof. Novak for giving me the time and allowing me to do rotations in their labs. I would also like to thank Dr. Gralnick for including me in his lab for over two years previous to me starting graduate school where I learned a lot about both science and what it means to be a scientist. I would like to thank my classmates: Carol, Marissa and Mason. Completing graduate school was a lot easier because of you three. I would like to thank my friends and family for their support and for dealing with me, especially while writing this thesis. And most of all I would like to thank Andrea. Words cannot express how grateful I am to you.

Table of Contents

List of Tables

List of Figures

Chapter 1
Sequencing Sugar Producing Algae

INTRODUCTION

Microalgae have received intense interest from a biotechnology standpoint due to their rapid growth rates, an ability to be cultured in either fresh or brackish water, and for their potential to produce a range of bioproducts with commercial potential (Hu *et al.*, 2008). Many of the microalgae that fall within the subset of organisms commonly referred to as green algae are members of the division Chlorophyta, and are found throughout the globe in a broad range of ecosystems. These algae play a vital role in the global carbon cycle, using light energy to drive photosynthesis to fix atmospheric carbon dioxide into a variety of reduced carbon compounds, including lipids and carbohydrates. These carbohydrates accumulate intracellularly as either starch for energy storage or as components of the cell wall (Domozych *et al.*, 2012, Ho *et al.*, 2013).

In recent years, the collection of data relating to the genetic blueprints of many plants, bacteria, archaea, and fungi has grown at a rapid rate. Yet for the subset of plants that constitute the green algae, which represent a foundational cornerstone in our understanding of photoautotrophs, we currently have a shockingly small number of published genomes. Only 31 green algal genomes have been published in NCBI as of this writing. In this work, we have de novo assembled and annotated two strains of green algae to expand this number of genomes and also investigated the genomic and transcriptomic basis for extracellular sugar production by certain species of green algae.

It is well documented that many green algae strains form symbiotic associations with invertebrates like the hydroid polyp *Hydra viridis*, the heliozoan *Acanthoscystis turfacea*, the freshwater sponge *Spongilla fluviatilis* and ciliates such as *Paramecium bursaria* (Fischer *et al.*, 1989, Pröschold *et al.*, 2011). Several symbiotic microalgal species of the *Chlorella* clade have been isolated and characterized from *P. bursaria* (Hoshina *et al.*, 2010) where the algae are maintained within perialgal vacuoles inside the

host. The environmental conditions within the perialgal vacuoles in these various hosts are not fully understood or adequately characterized but the pH of these environments have been proposed to be approximately pH 5.0 in certain Hydra (Douglas & Smith, 1984). The ability of certain species of green algae to establish a stable symbiotic association with *P. bursaria* depends on several factors, including the transfer of simple fixed carbon compounds (referred to as photosynthates) from the algae to the host which consume them (Ayala & Weis, 1987). Maltose, glucose and glucose-6-phosphate have been found to be the predominant photosynthates excreted by symbiotic green algae in many of these associations (Kessler *et al.*, 1991). In some cases, these symbiotic algae have developed such close associations with their hosts that they are unable to survive outside of the host relationship (Bosch, 2012). These specific associations are of interest for developing an understanding of the pathways and regulation of sugar production and export.

The release of sugars to the extracellular space in some strains of green algae isolated from symbiotic associations is well documented (Mews & Smith, 1982, Douglass & Smith, 1984, Matzke *et al.*, 1990, Kessler *et al.*, 1991, Dorling *et al.*, 1997, Summerer *et al.*, 2007, Shibata *et al.*, 2016). Symbiotic algae have been reported to release the maximum-level of sugars at slightly acidic pH values (~pH 5.7), but not release these sugars at more neutral pH values (~pH 7.6) (Kessler *et al.*, 1991). A study by the European Space Agency (Brechignac & Schiller, 1992) reported that endosymbiotic *Chlorella* sp. SAG 241.80, now named *Micractinium conductrix* SAG 241.80, produced a maximum of 5 g/L of maltose at pH 5.7 in batch culture. Some studies have correlated the acid tolerance to growth, maltose excretion and symbiotic ability in certain green algae (Kessler *et al.*, 1991, Brechignac & Schiller, 1992, Dorling *et al.*, 1997). However, the molecular mechanism of maltose excretion in relation to acid tolerance of growth and symbiotic ability still remains poorly understood for the algae in these associations, including the pathway (or pathways) involved in converting starch in the chloroplasts to sugars and transporting them outside of the cell (Shibata *et al.*, 2016). A sugar releasing pathway has been elucidated in *Arabidopsis thaliana* responsible for that organism's ability to secrete sucrose into the extracellular space, the difference being

that cytosolic maltose is converted to sucrose before it is transported out of the *A. thaliana* cell (Lu & Sharkey, 2006). Understanding this mechanism in algae could lead to the potential application of sugar production without the need for harvesting as the product is excreted straight into the media (Brechignac & Schiller, 1992, Ducat *et al.*, 2012).

C. sorokiniana was chosen to be the negative-control to be sequenced as it exhibits a very different phenotype when grown at the slightly acidic pH conditions when compared to *M. conductrix*. Selection of an optimal green alga strain as a negative-control for extracellular sugar production is limited in part based on the small numbers of algal genomes that are available and a very limited understanding of the strains that produce extracellular sugars. Thus, *C. sorokiniana* UTEX 1602 was selected from within our own laboratory stocks as we had determined that it does not produce appreciable amounts of extracellular sugars under similar condition to what is found for *M. conductrix* in preliminary studies, and *C. sorokiniana* is commonly used as a model species (Hu *et al.*, 2008, Barney *et al.*, 2015, Concas *et al.*, 2016, Li *et al.*, 2016, Wang *et al.*, 2016). In addition, *C. sorokiniana* UTEX 1602 is also a model for symbiotic associations with bacteria (de-Bashan *et al.*, 2002).

In this work, we conducted a comparative phenotypic study using a non-obligate, symbiotic, sugar-releasing green algae *Micractinium conductrix* SAG 241.80 (previously named *Chlorella sp.* SAG 241.80) and a non-symbiotic green algae strain *Chlorella sorokiniana* UTEX 1602 that lacks the sugar releasing phenotype. Our results confirmed the *M. conductrix* maltose excretion phenotype under slightly acidic conditions (pH 5.7), while the excretion of simple sugars was noticeably absent in *C. sorokiniana* under the same conditions. To investigate the potential pathways involved in maltose release at a lowered pH in *M. conductrix*, we used PacBio and Illumina sequencing technology to assemble and annotate the draft genomes of both strains of green algae. We compared the transcription of both the *M. conductrix* SAG 241.80 and *C. sorokiniana* UTEX 1602 strains at two pH conditions that result in different sugar excretion phenotypes using RNA-Seq to determine their metabolic response to the pH homeostasis. Herein, a genomic and transcriptomic characterization of *M. conductrix* SAG 241.80 and *C.*

sorokiniana UTEX 1602 with regard to their pH homeostasis and sugar excretion is presented, along with a proposed pathway to cytosolic sugar accumulation and a discussion of potential genes and mechanisms that may play a role in extracellular maltose release in *M. conductrix*.

MATERIALS AND METHODS

M. conductrix Media

Micractinium conductrix sp. SAG 241.80 (previously named *Chlorella sp.* SAG 241.80) was obtained from the Algal Culture Collection of the Plant Physiology Institute at the University of Goettingen, Germany. The algae was maintained on solid freshwater SAG medium plates (1.4 mM K_2HPO_4, 0.3 mM $MgSO_4 \cdot 7H_2O$, 0.17 mM $CaCl_2 \cdot 2H_2O$, 0.43 mM NaCl, 1.5 mM Na_2SO_4, 27 µM $(NH_4)_5Fe(C_6H_4O_7)_2$, 7.1 mM $NaNO_3$, 12 mM $Na_2S_2O_3 \cdot 5H_2O$, 100 µL vitamin solution [50 mg thiamine HCl, 1 mg biotin and 1 mg cyanocobalamin in 10 mL dH_2O], 2 mL Basal Medium Eagle (BME) vitamin solution [P/N B6891, Sigma Aldrich, St. Louis, MO], 1 mL trace element solution [1 g boric acid, 1 g sodium EDTA, 200 mg $MnCl_2 \cdot 4H_2O$, 20 mg $ZnCl_2$, 15 mg $CuCl_2 \cdot 2H_2O$, 15 mg $Na_2MoO_4.2H_2O$, 15 mg $CoCl_2 \cdot 6H_2O$ and 10 mg KBr in 1 L dH_2O] and 12 g of plant grade agarose [P/N A7921, Sigma Aldrich], pH 7.6) at 22 °C. Modified freshwater SAG medium was used as the liquid media for experiments (1.4 mM K_2HPO_4, 0.3 mM $MgSO_4 \cdot 7H_2O$, 0.17 mM $CaCl_2 \cdot 2H_2O$, 0.43 mM NaCl, 1.5 mM Na_2SO_4, 30 µM $(NH_4)_5Fe(C_6H_4O_7)_2$, 8.2 mM $NaNO_3$, 4 mM $Na_2S_2O_3 \cdot 5H_2O$, 50 µL polypropylene glycol, 200 µL vitamin solution, 2 mL BME vitamin solution and 3 mL trace element solution into 1.4 L dH_2O, pH 7.6). For low pH experiments (pH 5.7) 10 mM MES (2-(*N*-morpholino)ethansulfonic acid) was added as buffer.

C. sorokiniana Media

Chlorella sorokiniana UTEX 1602 was obtained from the Culture Collection of Algae at the University of Texas, Austin. The algae was maintained on agar plates at 22 °C that are a derivative of Bold's Basal medium as described previously (Villa *et al.*, 2014). Modified freshwater medium was used as the liquid media for experiments (1.4

mM K_2HPO_4, 0.3 mM $MgSO_4 \cdot 7H_2O$, 0.17 mM $CaCl_2 \cdot 2H_2O$, 0.43 mM NaCl, 1.5 mM Na_2SO_4, 44 μM $(NH_4)_5Fe(C_6H_4O_7)_2$, 8.2 mM $NaNO_3$, 50 μL polypropylene glycol and 1 mL trace element solution into 1.4 L dH_2O, pH 7.6), with slightly different vitamin and trace element concentrations than the modified freshwater SAG media. For low pH experiments (pH 5.7) 10 mM MES (2-(*N*-morpholino)ethansulfonic acid) was added as buffer.

Algal Growths in Tubular Photobioreactors

Algal cultures were grown (N=2) in glass tubular photobioreactors with an internal volume of 1.5 L (5 cm inner diameter by 80 cm length) containing a long capillary tube (1 mm inner diameter by 75 cm length) which was used to supply 0.2% CO_2 enriched air at a constant flow rate of 0.3 L gas per L of solution per minute. The aeration served as a source of CO_2, provided oxygen exchange and was the sole source of mixing to the culture during the experiments. *M. conductrix* and *C. sorokiniana* cells were grown on agar plates at room temperature on LED light beds. Once a sufficient mass of cells had grown, the cells were scraped with a sterile loop and inoculated into their respective liquid media. Cultures were grown at 22 °C with a 14-hour light and 10-hour dark cycle, to mimic a natural light cycle. Two, 1 mL samples were drawn from each tube daily for 10 days. One sample was used for measuring optical density (OD) at wavelength (λ) 750 nm. The second sample was centrifuged (2 minutes at 20000 g) and the supernatant was saved and frozen (-20 °C) for future analysis. Both algae were grown at a standard pH (7.6) and a low pH (5.7)

Sugar Concentration Analysis

Enzymatic assays were used to quantify the concentration of both glucose and maltose excreted by the algae into the media. Glucose was determined by putting 200 μL of the thawed supernatant sample into a plastic cuvette and mixing in 1 mL of the glucose assay reagent from the sucrose assay kit (Sigma Aldrich, St. Louis, MO) without adding invertase. After a 15-minute incubation period at room temperature, the OD at λ 340 nM is measured and compared to a standard curve to determine the concentration of glucose that was in the sample. The supernatant would be diluted if the OD reading was outside

the range of the standard curve. Maltose was determined by first mixing the 200 µL of supernatant with 100 mM potassium phosphate buffer pH 6.6 with 100 mM KCl and α-glucosidase (Sigma Chemical Company P/N G5003) and incubating at 37 °C for 30 minutes. The 1 mL of glucose assay reagent is then added without invertase, the sample is incubated for 15 minutes at room temperature and the OD is measured.

Genomic DNA Isolation

M. conductrix and *C. sorokiniana* cells were scraped with a sterile loop from agar plates and the genomic DNA was isolated using the ZR Fungal/Bacterial DNA Microprep kit as directed by the manufacturer (Zymo Research, Irvine, CA). Following isolation, the DNA quantity and quality was evaluated using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA).

Genome Sequencing

Genomic DNA of *M. conductrix* and *C. sorokiniana* was sequenced using the Illumina Hi-Seq 2500 (Illumina, San Diego, CA) at the University of Minnesota Genomics Center using the standard protocol for paired-end reads. Both strains were also sequenced using Pacific Biosciences (PacBio) Single Molecule Real Time (SMRT) technology at the Rochester Mayo Medical Genome Facility. 5 µg of raw genomic DNA was analyzed for size, quality and concentration by the Qubit dsDNA BR Assay (Invitrogen, Carlsbad, CA) and Fragment Analysis (Advance Analytical Technologies, Inc., Alkeny, IA) using the High Sensitivity Large Fragment Analysis kit. SMRTbell template libraries were generated using the SMRTbell Template Prep Kit 1.0. All purification and concentration steps were performed using AMPure PB beads (Pacific Biosciences, Menlo Park, CA). BluePippin Size Selection (Sage Science, Beverly, MA) was performed to optimize large insert libraries for improved long-reads. The annealing and binding reactions of BluePippin Size-Selected SMRTbell templates were carried out using the DNA/Polymerase Binding Kit P6 (Pacific Biosciences, Menlo Park, CA). The Annealing and Binding Calculator (version 2.3.1.1) provided by Pacific Biosciences was used to calculate the concentration of bound complex to be loaded onto the sample plate and to determine which MagBead Kit (Pacific Biosciences, Menlo Park, CA) loading

steps and recipes were ideal. DNA samples were run on the PacBio RSII SMRT DNA Sequencer (Pacific Biosciences, Menlo Park, CA) using the DNA Sequencing Reagent 4.0 kit and SMRT Cell v3 cells. Instrument parameters were setup as follows: MagBead collection protocol, 20,000 bp insert size, stage start and 240-minute movie collection time.

Genome Assembly

Sequence analysis was performed using a combination SMRT Portal v2.3.0, SMRT Analysis v2.3.0, SMRT Pipe v2.3.0 and SMRT View version 2.3.0. Default parameters in the HGAP 3 pipeline (Chin *et al.*, 2013) were used for the entire process, with the exception of the estimated genome sizes. Several iterations of the HGAP 3 procedure were conducted with gradually refined genome sizes until the estimated genome size entered and the assembled genome size were within 500 KB of each other. The initial assemblies were subjected to short-read polishing with the default Pilon pipeline (Walker *et al.*, 2014) to correct most remaining single nucleotide polymorphisms (SNPs) and insertion/deletion (indel) errors.

Genome Annotation

Both genomes were structurally annotated using the Augustus RNA-Seq integration pipeline, which uses Tophat to generate intron hits (Stanke *et al.*, 2008) as described by Bioinformatics Greifswald[1] to determine predicted genes. The default parameter set used *Chlamydomonas reinhardtii* as the reference species. RepeatScout (Price *et al.*, 2005) was used to identify repeated sequences with a k-mer size of 13. The predicted gene products were run through BLASTp (Altschul *et al.*, 1997) and the results were processed with Blast2Go to map and functionally annotate the predicted coding sequences using the default parameters (Conesa *et al.*, 2005)

Genome Comparison

The structurally annotated *M. conductrix* SAG 241.80 and *C. sorokiniana* UTEX 1602 genomes, and other available algal genomes (*Ostreococcus tauri*, *Micromonas sp.*

[1] http://bioinf.uni-greifswald.de/bioinf/wiki/pmwiki.php?n=IncorporatingRNAseq.Tophat

RCC299, *Coccomyxa subellipsoidea* C-169, *Volvox carteri*, *Chlamydomonas reinhardtii*, *Chlorella variabilis* NC64A and *Chlorella pyrenoidosa* FACHB-9) were analyzed by GET_HOMOLOGUES (Vinuesa and Contreras-Moreira, 2015), using all default parameters as described[2]. In summary, the amino acid sequences of annotated genes were used as BLASTp queries against each other, then OrthoMCL (Li *et al.*, 2003) was used to cluster homologous proteins into groups/families. A presence/absence pan-genome matrix was constructed based on the clustering result. A Venn Diagram was plotted in R using the VennDiagram package (Chen & Boutros, 2011) with the designated genomes. Graphics of the contigs were also generated in R.

Organellar Genome Annotation

Known chloroplast and mitochondrial genes were used to search for the organellar genomes among the assembled contigs obtained from the initial sequencing assembly using the BLAST search algorithm. The identified organellar contigs were manually inspected and cut in order to circularize them. The chloroplast and mitochondrial contigs were passed through ORFfinder (Wheeler *et al.*, 2003) to determine the possible open reading frames (ORF). The results were then run through BLASTp. Mapping and annotation were done with Blast2Go using default parameters. Mitochondrial and chloroplast genome graphics were produced using the software program GenomeVx (Conant and Wolfe, 2008).

RNA Isolation

M. conductrix and *C. sorokiniana* cells were collected in triplicate in both the low pH and standard pH cultures on days 7, 8 and 10 for *M. conductrix* and days 4, 5 and 8 for *C. sorokiniana* at approximately 2:00 PM (after at least six hours of continuous light) during exponential or early stationary phases of growth and frozen in liquid nitrogen for total RNA preparation. Briefly, cells were harvested by centrifugation at 14000 g for 1 minute and cell pellets were frozen in liquid nitrogen then stored at -80˚C. Frozen pellets were suspended in 1 mL Trizol and transferred to a bashing beads tube (ZR

[2] http://eead-csic-compbio.github.io/get_homologues/manual-est/

Fungal/Bacterial DNA Microprep kit, Zymo Research, Irvine, CA). The solution was vortexed with the beads for 2 minutes, then incubated at room temperature for 5 minutes. Next, 200 µL of chloroform was added to the mixture, vortexed well and centrifuged (12000 g for 3 minutes). The upper layer was carefully collected in an RNase-free tube and mixed with an equivalent volume of 95% ethanol (Sigma) and vortexed well. The solution was further purified using the Direct-zol RNA kit (Zymo Research, Irvine, CA) as directed by the manufacturer. The total RNA was collected and treated with RNase free DNase I (Zymo Research, Irvine, CA) for 10 minutes and then purified again using the Direct-zol RNA kit. Total RNA quantity and quality were measured using the NanoDrop 2000 spectrophotometer.

RNA Sequencing

Poly(A)$^+$ tag based mRNA isolation and RNA sequencing were performed at the University of Minnesota Genomics Center. Poly(A)$^+$ RNA was isolated using oligo(dT) magnetic beads and isolated mRNA was fragmented into small pieces at elevated temperature. The fragmented mRNA was used as template for cDNA synthesis using random primers. Further, 5' and 3' adapters were ligated to the mRNA and the adapter-ligated RNA was reverse-transcribed and amplified using PCR. The final PCR products were size fractioned and used for cDNA sequencing library construction. The cDNA sequencing libraries were sequenced using the HiSeq 2500 Illumina sequencing technique.

Transcriptome Analysis

Raw paired-end RNA-Seq reads in FASTQ format were analyzed using Gopher-pipeline, RNAseq-Pipeline Version 1.4[3], which assesses base call quality, cycle uniformity and contamination using fastQC[4]. It also maps reads to the reference genomes of *M. conductrix* and *C. sorokiniana* via HiSAT2 (Kim *et al.*, 2015) using our curated annotation, and then cleans the alignments using picard tools[5], and quantifies the

[3] https://bitbucket.org/jgarbe/gopher-pipelines

[4] http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[5] https://broadinstitute.github.io/picard/

Fragments mapped Per Kilobase of transcript per Million mapped reads (FPKM) using Cufflinks package (Trapnell *et al.*, 2010). The resulting FPKM table was imported into Cuffdiff to test for the differential expression using default parameters.

Assembly Quality Control

Three levels of quality control were run for both genome assemblies. Original Illumina DNA reads were aligned to the draft genome assemblies of both organisms using BWA-MEM (Li & Durbin, 2010)**.** RNA-Seq cDNA libraries were also aligned to the draft genome assemblies of both organisms HISAT2 (Kim *et al.*, 2015)**.** Both draft genomes were analyzed with BUSCO (Simão *et al.*, 2015) using the eukaryotic gene set.

Accession Numbers

Whole genome sequences of *Micractinium conductrix* SAG 241.80 and *Chlorella sorokiniana* UTEX 1602 are available in the NCBI GenBank data libraries under accession numbers LHPF00000000 (*M. conductrix* SAG 241.80 PRJNA290385) and LHPG00000000 (*C. sorokiniana* UTEX 1602 PRJNA290386). Transcriptional data is available through accession number GSE98781.

RESULTS

Confirmation of Sugar Excretion in *M. conductrix* and *C. sorokiniana*

As the first step in our work, we sought to confirm the release of simple sugars by *M. conductrix* and the lack of release by *C. sorokiniana*. The results obtained illustrated the stark difference in the phenotypes for *M. conductrix* at the two pH conditions tested (5.7 and 7.6) as well as the difference between the two species (Figure 1.1). Enzymatic assays confirmed elevated levels of maltose accumulating in the extracellular space for *M. conductrix* at pH 5.7, gradually increasing up to 700 μM maltose after 8 days. At pH 7.6, maltose levels were much lower (< 50 μM maltose) and stayed constant over the length of the growth. A similar trend was seen with glucose release, albeit at lower levels, getting to 100 μM glucose at day 8. For *C. sorokiniana* grown under identical conditions,

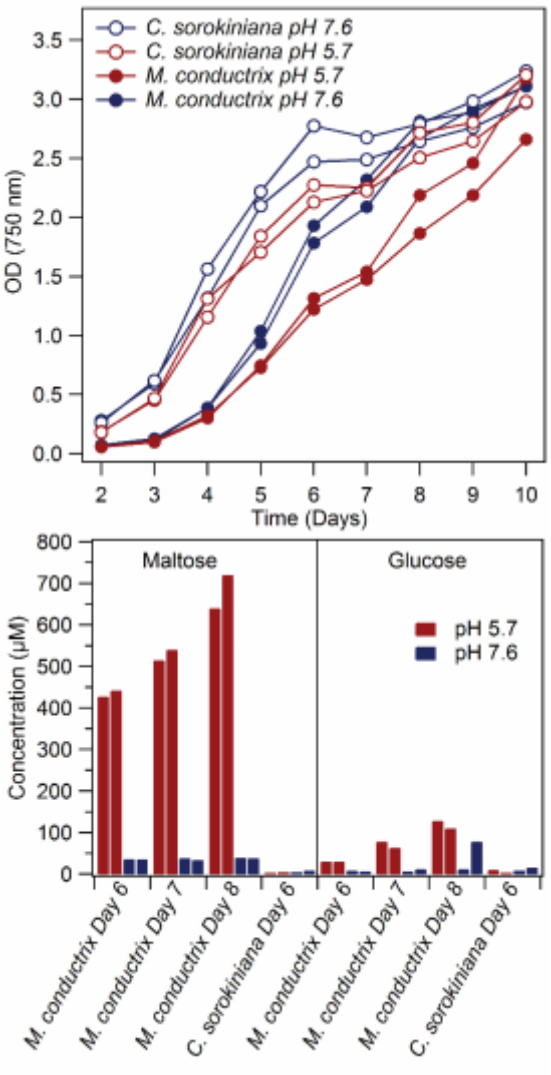


Figure 1.1. Culture growth and sugar accumulation by *Micractinium conductrix* SAG 241.80 and *Chlorella sorokiniana* UTEX 1602. Upper panel shows growth of replicate samples grown at two pH conditions (pH 5.7 and pH 7.6) for the two strains sequenced in this work. Lower panel shows the amount of maltose (left) and the amount of glucose (right) found for specific days under each pH condition. Three consecutive days are shown for *M. conductrix* SAG 241.80, while a representative sample is shown for *C. sorokiniana* UTEX 1602.

we found little evidence of any maltose or glucose excretion. Similar results were found when analyzing the sugar excretion levels by HPLC (results not shown).

Genomic features of *M. conductrix* and *C. sorokiniana*

M. conductrix and *C. sorokiniana* were both sequenced with PacBio SMRT sequencing to obtain large contigs and further polished by Illumina sequencing. While this approach did result in the assembly of a series of large contigs, it did not completely close the genomes for either species. The *M. conductrix* genome was sequenced with 57x coverage and the *C. sorokiniana* genome was sequenced with 56x coverage based on the PacBio assembly. A summary of the completed draft genomes for each organism is shown in Table 1.1, including a summary of the previously sequenced *Chlorella variabilis* NC64A genome (Blanc *et al*., 2010).

The *M. conductrix* and *C. sorokiniana* draft genomes were both larger than what was reported for *C. variabilis* (32% and 29% larger respectively). However, the overall gene counts for all three strains were relatively similar with all three species falling

Table 1.1. Genome Sequence information for *M. conductrix* SAG 241.80, *C. sorokiniana* UTEX 1602 and *C. variabilis* NC64A

	M. conductrix SAG 241.80	*C. sorokiniana* UTEX 1602	*C. variabilis* NC64A
Nuclear genome			
Genome Size (Mb)	60.9	59.4	46.2
Chromosomes	≥13	≥13	12
GC content	67.36%	64.09%	67%
Gene Count	9349	9587	9791
Avg protein length (aa)	637	631	456
Avg # of exons per gene	13.9	14.6	7.3
Avg exon length (bp)	152	146	170
Avg intron length (bp)	250	231	209
Coding sequence	29.36%	30.55%	29
Chloroplast genome			
Genome Size (bp)	129,445	104,485	124,793
Gene Count	118	115	114
GC content (%)	34.8%	33.7%	33.9%
Mitochondrial genome			
Genome Size (bp)	75,014	56,631	78,500
Gene Count	56	52	62
GC content (%)	29.4%	30.3%	28.2%

between 9349 and 9791 genes. *M. conductrix* and *C. variabilis* had comparable nuclear genome GC contents (~67%) while *C. sorokiniana* was slightly lower at 64.09%. Chloroplast and mitochondrial genomes were also assembled for *M. conductrix* and *C. sorokiniana* (Table 1.1). The GC content was much lower for the organellar genomes compared to the nuclear genome. The *M. conductrix* and *C. variabilis* (Orsini *et al.*, 2016) chloroplast and mitochondrial genomes were similar in size, while the *C. sorokiniana* organellar genomes were smaller in comparison.

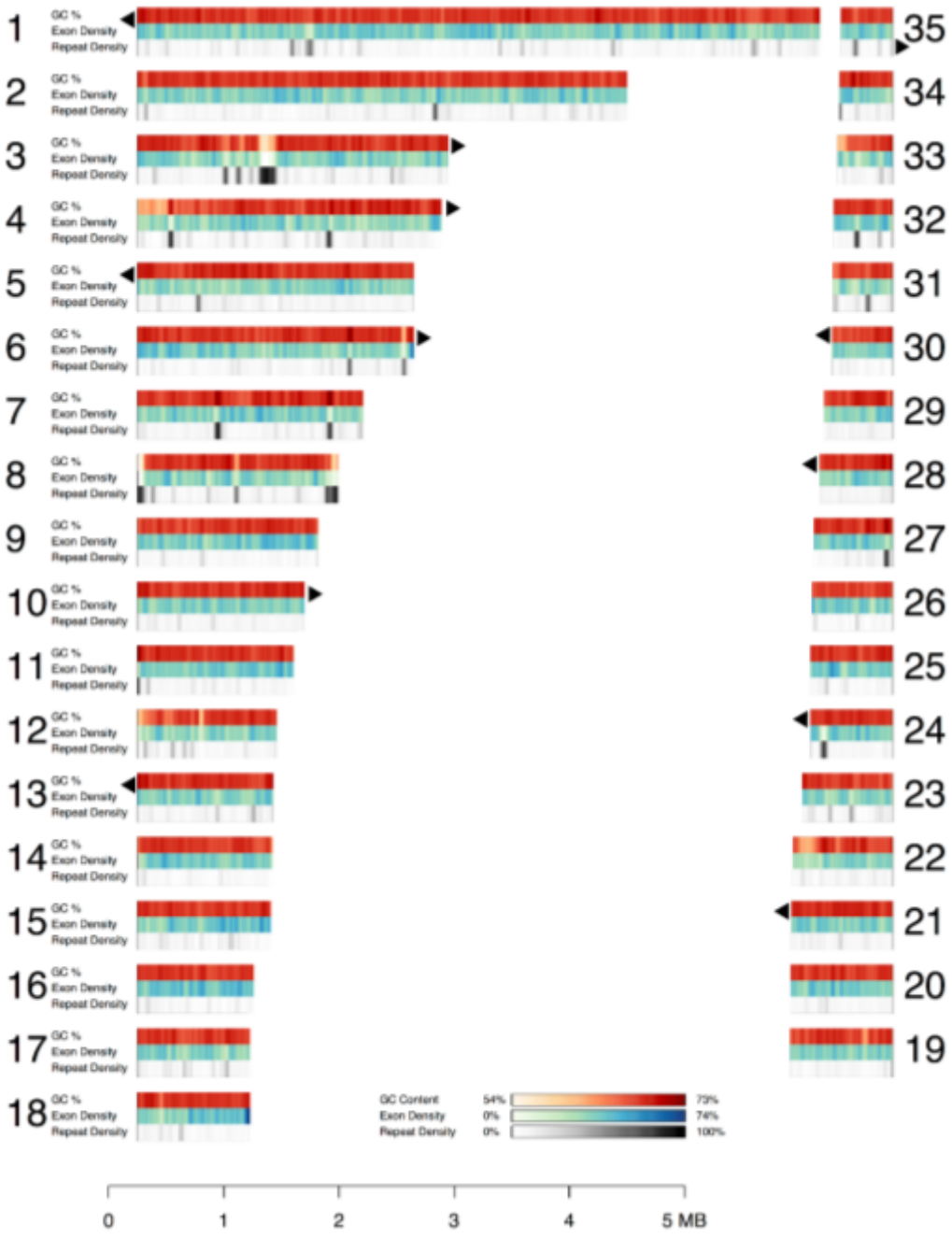
One reason why the nuclear genome sequence did not fully close is that the DNA is likely organized within distinct chromosomes, as was reported for *C. variabilis* NC64A and multiple additional Chlorophytes (Blanc *et al.*, 2010). The *C. variabilis* sequencing using whole-genome shotgun Sanger sequencing resulted in several hundred scaffolds greater than 1 kb, the largest being just over 3 Mb, with 89 percent of the genome represented within just 30 scaffolds. The *M. conductrix* genome sequencing here resulted in several longer scaffolds, with the two largest being 5.96 Mb and 4.28 Mb in size, with 80 percent of the genome represented by just 35 major contigs (Figure 1.2a). An analysis of the 299 assembled contigs searching for the telomere repeating sequence $(TTTAGGG)_n$ revealed at least 26 contigs containing arrays of this repeat on one end of the scaffold, indicating the likely end of a chromosome, which would indicate at least 13 chromosomes. The *C. sorokiniana* UTEX 1602 genome sequencing resulted in seven contigs greater than 3 Mb, the largest being 4.19 Mb, with 80 percent of the genome represented by just 17 major contigs (Figure 1.2b). An analysis of 158 assembled contigs searching for the telomere repeating sequence revealed at least 24 contigs containing potential telomeres, although one of those contigs had potential telomeres on both sides, indicating at least 13 chromosomes. The contig containing telomere sequences on both ends, contig_2, also had a strong repeated sequence in the center that might represent a centromere, indicating that this contig most likely represents a complete chromosome.

Genome Assembly Quality Control

We performed multiple levels of quality control for the two genome assemblies (Table 1.2). We first aligned the Illumina DNA reads with the genome assemblies for both organisms, of which 98.92% of the reads mapped back to *Micractinium conductrix* SAG 241.80 with a mapping error rate of 0.34%, and 99.29% of the reads mapped back to *Chlorella sorokiniana* UTEX 1602 with a mapping error rate of 0.92%. We also aligned the RNA-Seq cDNA libraries to the assembled genomes, 94.57% of which correctly mapped back to *M. conductrix* and 95.03% mapped back to *C. sorokiniana*.

Figure 1.2. Graphical depiction of the largest contigs assembled in the draft genomes of (A) *Micractinium conductrix* SAG 241.80 and (B) *Chlorella sorokiniana* UTEX 1602. Likely telomeres are represented by triangles (▲) at the corresponding ends of the contigs.

A)



B)

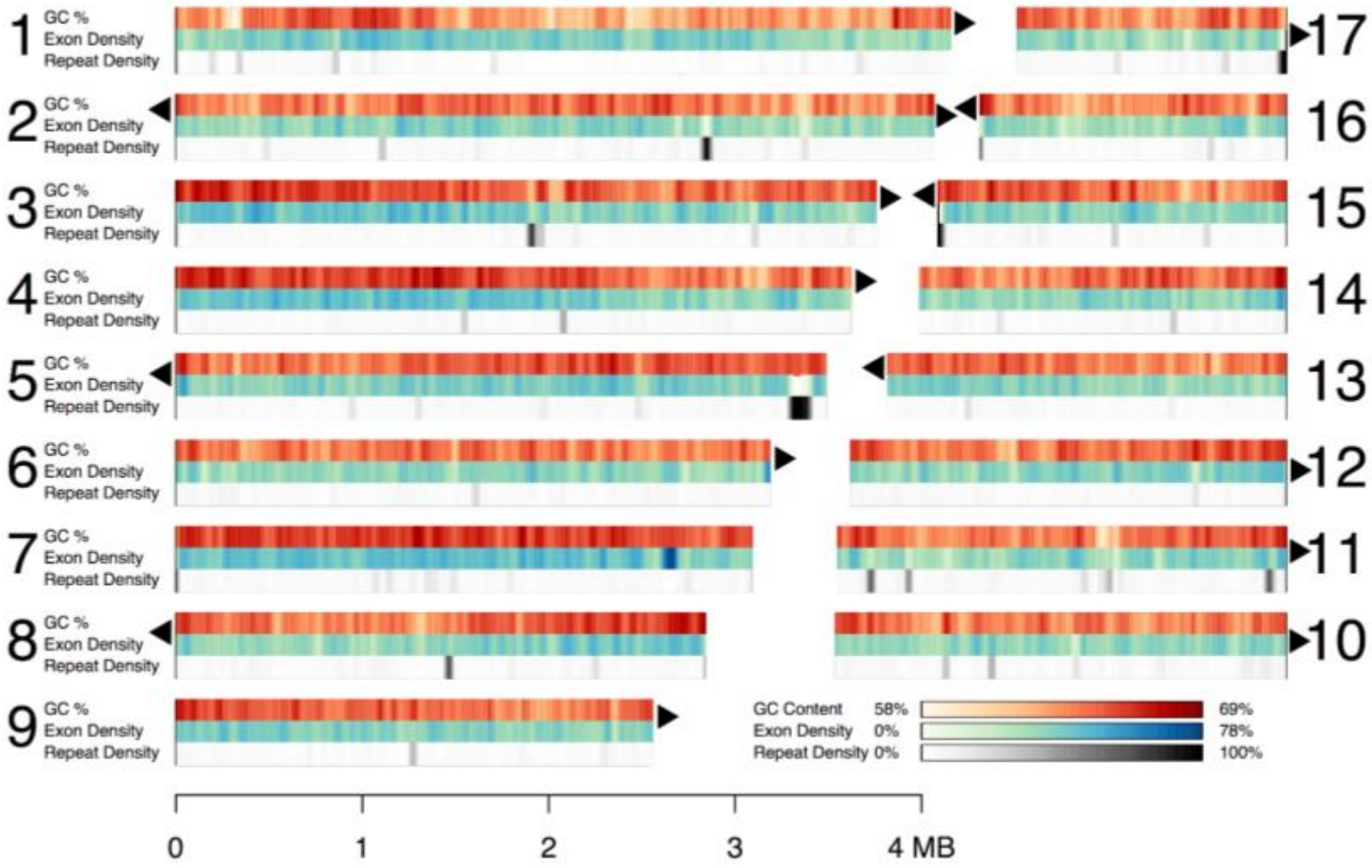

1
GC %
Exon Density
Repeat Density
17
2
16
3
15
4
14
5
13
6
12
7
11
8
10
9
GC Content 58% 69%
Exon Density 0% 78%
Repeat Density 0% 100%
0
1
2
3
4 MB

Table 1.2. Genome assembly quality for *M. conductrix* SAG 241.80 and *C. sorokiniana* UTEX 1602

	M. conductrix SAG 241.80	*C. sorokiniana* UTEX 1602
Number of Contigs	301	160
Genome Size (Bb)	61,020,757	59,568,085
Largest Contig (Bp)	5,958,903	4,192,086
N50 (Bp)	1,210,495	2,592,956
Coverage	56.84 x	55.87 x
Illumina Mapped Reads	98.92%	99.29%
Illumina Properly Paired	97.33%	none
Illumina Error Rate	0.34%	0.92%
RNA-Seq Mapped Reads	94.57%	95.03%
BUSCO Results (Eukaryotic Gene Set)		
Complete Single-Copy	80.2%	78.9%
Complete Duplicated	1.7%	0.7%
Fragmented	1.0%	1.0%
Missing	17.1%	19.4%

Our final level of quality control was to run BUSCO (Simão, *et al.*, 2015) to look for orthologs for the universal single-copy (USCO) genes found in eukaryotes, as provided by BUSCO. For *M. conductrix*, 81.9% of the eukaryotic USCO genes were present and complete while in *C. sorokiniana*, 79.6% of USCO genes were present and complete which is comparable to *C. variabilis* NC64A (Roth *et al.*, 2017).

Phylogeny

Micractinium conductrix SAG 241.80 was formerly referred to as a *Chlorella* species (Kessler *et al.*, 1991), but was recently renamed under the genus of *Micractinium* (Proschöld, *et al.*, 2011). While some have called to rename members of this species as *Micractinium reisseri* (Hoshina *et al.*, 2010), we have selected to adopt the naming designation of *Micractinium conductrix* SAG 241.80, as this is the name currently referenced by the culture collection from which our strain was obtained. *M. conductrix* SAG 241.80 falls within the class of Trebouxiophyceae of the division Chlorophyta.

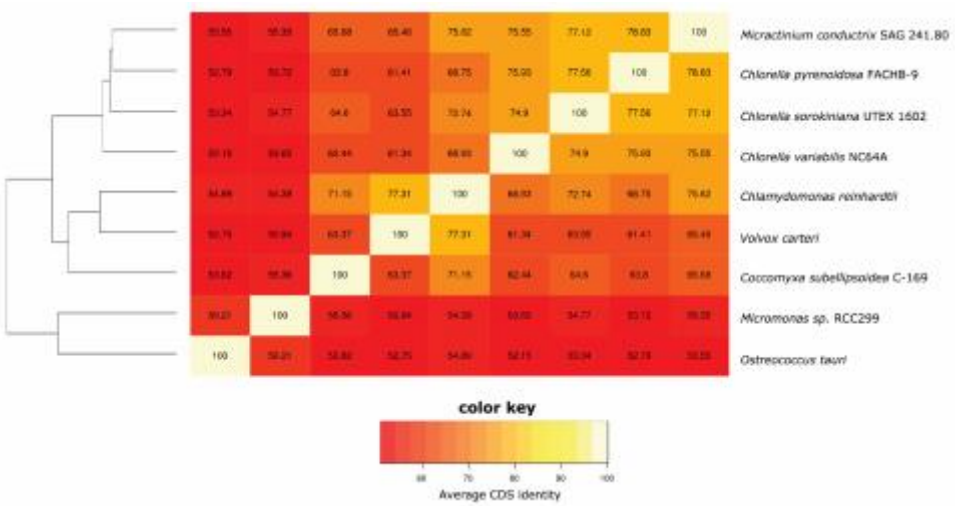


Figure 1.3. Hierarchical clustering based on the coding DNA sequences (CDS) of members of the *Chlorella* clade and other select green algae. Similarity represented via heatmap

---

Previous phylogenetic trees have been constructed using maximum likelihood based on small subunit rRNA and ITS2 sequences, showing that *C. variabilis* NC64A and *M. conductrix* SAG 241.80 both fall within the *Chlorella* clade (Hoshina *et al.*, 2010). We constructed a heat map and a phylogenetic tree based on the coding DNA sequence (CDS) translated amino acid similarities of sequenced members of the *Chlorella* clade along with a few other select green algae (Figure 1.3). The average nucleotide identity score was calculated based on BLASTp comparison among the 285 homologous genes that are common in all nine species.

The two most closely related strains to the strains that we sequenced were *Chlorella pyrenoidosa* FACHB-9, which is a free-living strain, and *Chlorella variabilis* NC64A, which was isolated from a *Paramecium* host. Neither strain has been shown to produce extracellular sugars. Surprisingly for these members of the *Chlorella* clade, the two strains isolated from *Paramecium* (*M. conductrix* and *C. variabilis*) are the least related as the free-living strains fall between them on the phylogenetic tree.

After annotating the genomes of *M. conductrix* SAG 241.80 and *C. sorokiniana* UTEX 1602 we organized the genes into groups of unique and shared genes and visualized this as a Venn Diagram (Figure 1.4), including other sequenced members of

the *Chlorella* clade. There were 6084 unique genes in *M. conductrix* (67.8% of its genes), 6480 unique genes in *C. sorokiniana* (69.2%), and 2886 genes were found in both strains. The large number of unique genes was not helpful for narrowing down possible genes that could cause the difference in the sugar excreting phenotype.

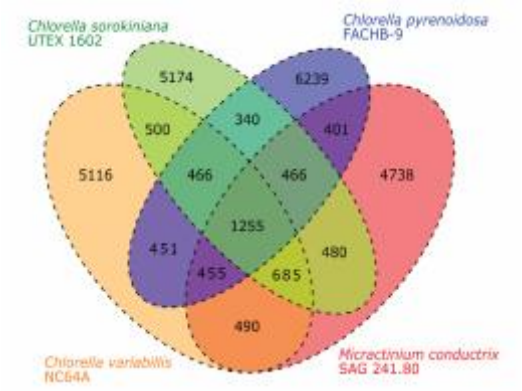


Figure 1.4. A Venn diagram representation of shared and unique protein families of *M. conductrix* SAG 241.80, *C. sorokiniana* UTEX 1602, *C. variabilis* NC64A and *C. pyrenoidosa* FACHB-9. Generated using the VennDiagram package in R.

Putative Pathway from Chloroplastic Starch to Cytosolic Maltose

In an attempt to construct a simple putative pathway for how the *M. conductrix* produces extracellular sugars, we have looked to identify homologs of key genes whose products are involved in the sucrose secretion pathway previously reported in *A. thaliana* (Lu & Sharkey, 2006) in the newly sequenced algae genome (Figure 1.5).

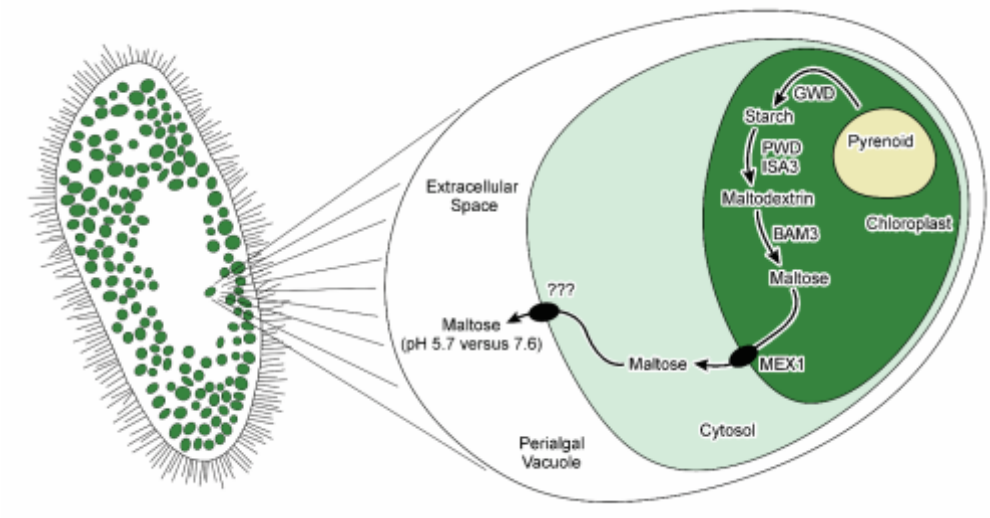


Figure 1.5. Illustration of the symbiotic relationship between *Paramecium bursaria* and *Micractinium conductrix* SAG 241.80 and the potential pathway resulting in maltose release by *M. conductrix* based on the sucrose secretion pathway in *Arabidopsis thaliana* (Lu & Sharkey, 2006). The ciliate *P. bursaria* harbors several hundred cells of symbiotic algae within perialgal vacuoles (enlarged depiction of perialgal vacuole containing a single *M. conductrix* cell on right). The potential maltose production pathway and key genes that are required or proposed for this pathway show how carbon is shuttled from the pyrenoid to *P. bursaria*.

Table 1.3. Potential genes constituting the maltose excretion pathway from *M. conductrix* SAG 241.80, *C. sorokiniana* UTEX 1602 and *C. variabilis* NC64A

Enzyme	*A. thaliana* Query Gene	*M. conductrix* SAG 241.80	*C. sorokiniana* UTEX 1602	*C. variabilis* NC64A
Glucan water dikinase (GWD)	At1g10760	g72	g8031	37464, 135335, 21770
Phosphoglucan water dikinase (PWD)	At5g26570	g7307, g72, g3110[A], g3535	g3326, g8031, g5801[A]	21770, 37464, 54374, 57278, 54232
Isoamylase (ISA3)	At4g09020	g495, g6850, g3928[A], g4952[A], g7390	g2393, g1338, g4700	34754, 58311, 142212
β-amylase (BAM3)	At4g17090	g4397, g7327	g3868, g3347	134683, 139821, 134682, 143931
Maltose Transporter (MEX1)	At5g17520	g9065	g6937	138472

Genes identified as potential homologs were determined using BLASTP to compare the amino acid sequences of the translated gene products with a cutoff E-value of 10^{-10} utilizing the BLOSUM62 matrix. The results are listed in order of ascending E-values (most related gene listed first). Gene transcription was measured in triplicate and tested at pH 5.7 and 7.6 as part of the transcriptional studies. None of the listed genes showed a significant difference in transcription (> 1.5-fold change in either direction, $p < 0.05$).

[A] Genes that had annotations different than the query gene name. These genes tend to have higher E-values and are most likely related due to similarity within some domains rather than similarity throughout the entire gene.

An analysis of the genes identified from the genomes sequenced in this study (*M. conductrix* SAG 241.80 and *C. sorokiniana* UTEX 1602) and a previously reported genome from a strain which was also isolated from a *Paramecium* (*C. variabilis* NC64A) found homologs for all of the genes described in the pathway in *A. thaliana* leading from chloroplastic starch to cytosolic maltose in all three algal species (Table 1.3). From these results, we can presume that there is no requirement for additional genes to facilitate the pathway to the point of cytosolic maltose, though it is not possible to definitively state that additional pathways do not exist.

Table 1.4. Potential sugar transport genes in *M. conductrix* SAG 241.80, *C. sorokiniana* UTEX 1602 and *C. variabilis* NC64A

Sugar Transporter	Query Gene	*M. conductrix* SAG 241.80	*C. sorokiniana* UTEX 1602	*C. variabilis* NC64A
MalFGK (*E. coli*)	Y75_RS21025-35	none	none	none
AtSWEET11 (*Arabidopsis*)	At3g48740	g3343[C], g5513[C]	g4186[C]	59697
SbSWEET13A (Sorghum)	Sb08g013620	g3343[C], g5513[C]	g4186[C]	59697
AtSUC1[A]	At1g71880	none	none	none
AtSUT4[A]	At1g09960	none	none	none
OsSUT1[A] (Rice)	Os03g0170900	none	none	none
BvTST2[B] (Beet)	XP_010678631	g1657, g2551, g1592, g859, g218, g1786[D], g2658 [2.0][E], g5839	g5784, g8497, g4825, g2863, g2750, g3243, g7686, g801, g7601, g6747, g8683, g477[D]	137750, 8066, 136418, 25635, 23239, 55965, 33950, 59832, 59732, 139619, 33524

Known sugar transporters from *Escherichia coli* (Hayashi *et al.*, 2006), *Arabidopsis thaliana* (Manck-Götzenberger & Requena, 2016), *Beta vulgaris* (Jung *et al.*, 2015), *Oryza sativa* (Scofield *et al.*, 2007, Eom *et al.*, 2011), and *Sorghum bicolor* (Bihmidine *et al.*, 2016) were used as queries to identify potential homologs using BLASTP to compare the amino acid sequences of the translated gene products with a cutoff E-value of 10^{-10} utilizing the BLOSUM62 matrix. The results are listed in order of ascending E-values. Gene transcription was measured in triplicate and tested at pH 5.7 and 7.6 as part of the transcriptional studies. The annotated gene names and E-values for this table can be found in supplemental tables S3 and S4.

[A] Also analyzed similar genes AtSUC2 (At1g22710), AtSUC5 (At1g71890), and OsSUT2 (Rice, Os12g44380) and obtained similar results (no homologs found).

[B] Also analyzed similar genes AtTMT1 (At1g20840), AtTMT2 (At4g35300), AtTMT3 (At3g51490) and AtVGT1(At3g03090) and obtained similar results (the same homologs found in a slightly different order of E-value).

[C] Genes that had annotations different than the query gene name. These genes tend to have higher E-values and are most likely related due to similarity within some domains rather than similarity throughout the entire gene.

[D] Genes that had a minimum of one zero value in the RNA-Seq data. These genes were not included in any transcriptional analysis.

[E] The only gene (*M. conductrix* g2658) that showed a significant difference in transcription (> 1.5-fold change in either direction, $p < 0.05$) with the fold increase shown in brackets [pH 5.7/pH 7.6].

Extracellular Maltose Transport

The transport of the cytosolic maltose to the extracellular space becomes problematic, as we are unaware of any report of a specific gene that has been identified which transports maltose specifically *outside* of the cell. Our understanding of the transport of sugars throughout the vascular systems in land plants is related primarily to studies of the disaccharide sucrose, genes coding proteins responsible for sucrose transport in plants are well established (Lu & Sharkey, 2006, Sauer, 2007, Ainsworth & Bush, 2011, Eom *et al.*, 2011, Jung *et al.*, 2015). The lack of a good candidate to search for maltose transporter homologs in the algae is an obstacle to defining a complete preliminary pathway in this species. A list of several classes of genes that have been found to play a role in the transport of sugars in different model plants is provided (Table 1.4), along with homologous genes from the three algae strains evaluated here. This includes SWEET family transporters that move glucose or sucrose (Chen *et al.*, 2010), Suc transporters that move sucrose (Sauer, 2007), and Tonoplast Sucrose Transporters (TSTs) that move mostly sucrose across cell walls. (Bihmidine *et al.*, 2016).

Transcriptional Changes

We performed RNA-Seq experiments on both *M. conductrix* and *C. sorokiniana* at both pH 5.7 and 7.6 to identify genes that are differentially transcribed under conditions of maltose release. Triplicate samples were taken, fragments mapped per kilobase of transcript per million mapped reads (FPKM) were calculated and fold change was determined. Any gene that registered a zero FPKM value in any sample for either strain was not included in any transcriptional analysis. We also performed a Welch's t-test to determine if these fold changes were statistically significant ($p \leq 0.05$).

None of the genes constituting the putative pathway to cytosolic maltose (Figure 1.5, Table 1.3) in either *M. conductrix* or *C. sorokiniana* showed any appreciable differences in transcription, defined here as a 1.5-fold change in either direction, when grown at pH 5.7 versus pH 7.6. The lack of change in transcript levels for these genes was unexpected, but it is possible that the proteins involved in the pathway are not rate

Table 1.5. Genes with elevated transcription levels in *M. conductrix* SAG 241.80 at pH 5.7 versus pH 7.6

Gene (amino acid size)	Fold Change [5.7/7.6]	P-value (Welch's t-test)	Gene Annotation
g1861 (69)	144.5	0.015	H-type lectin domain
g7102 (550)	13.6	0.029	NADH dehydrogenase
g6744 (333)	8.4	0.015	Hypothetical protein
g5932 (92)	6.0	0.029	Hypothetical protein
g7013 (602)	5.9	0.034	Expressed protein
g7141 (1514)	5.8	0.008	ATP-binding cassette transporter[A]
g7595 (1411)	4.8	0.009	Sodium hydrogen exchanger 7[A]
g6714 (518)	4.5	0.028	Ligand-gated ion channel[A]
g3233 (71)	4.4	0.048	H-type lectin domain
g3870 (123)	4.3	0.008	Chitinase, lysozyme, NLP P60
g5861 (2260)	4.1	0.006	Polysaccharide deacetylase
g6292 (759)	3.9	0.022	Concanavalin A-like lectin glucanase
g4962 (389)	3.7	0.007	RNA-binding Musashi homolog 1
g642 (882)	3.5	0.010	Glycoside hydrolase
g8691 (1181)	3.4	0.006	Small glutamine-rich tetratricopeptide repeat-containing

[A] Genes encoding potential transport associated proteins

limited at either pH so there might be no need to produce more protein to also produce an elevated amount of maltose from starch. For the genes that are homologous to the sugar transporters we queried (Table 1.4), only one gene had a fold change that was greater than our set threshold, g2658, which is showed a two-fold increase in transcription and is annotated as a (H)+ hexose cotransporter 2. None of the other transporter homologs showed any appreciable difference in transcription under the two pH conditions. One gene from both species was omitted from analysis due to having at least one zero FPKM value (g1786 in *M. conductrix* and g477 in *C. sorokiniana*).

We also looked at the genes in *M. conductrix* that did have an appreciable transcriptional increase when grown at the lowered pH 5.7. Only a small set of genes were found to have a more than 3-fold increase in transcription that were also found to be statistically significant (Table 1.5). Notably, this table includes several genes that might be of interest in relation to membrane transport, including an ATP-binding cassette transporter (g71141), a sodium hydrogen exchanger (g7595), and a Ligand-gated ion channel (g6714), in addition to several hypothetical proteins. Several of the genes also

contained likely lectin domains (g1861 and g3233) which are carbohydrate-binding, including the gene that showed the greatest different in levels of transcription (g1861 with a ~144-fold increase at pH 5.7 versus pH 7.6).

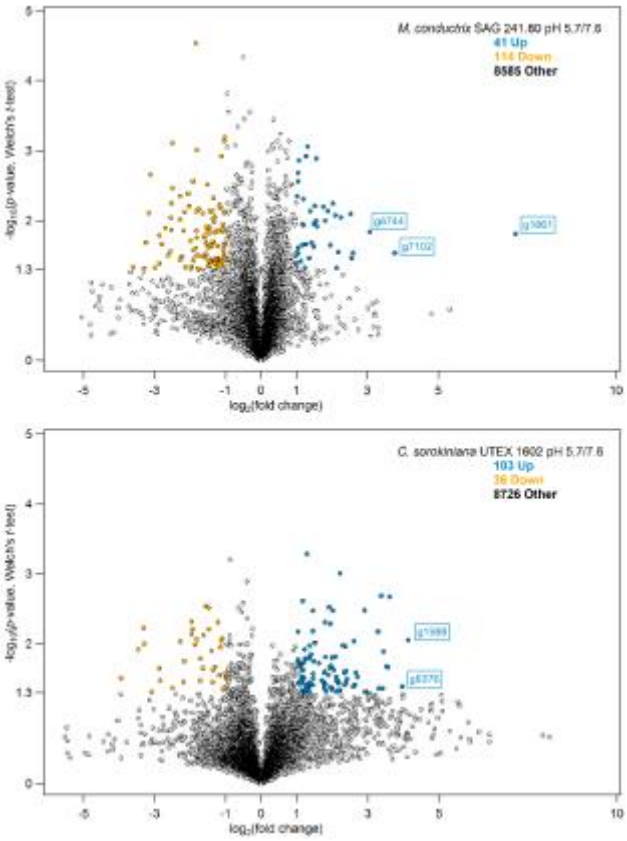


Figure 1.6. Volcano plots comparing the fold change in gene transcription between *M. conductrix* SAG 241.80 (top) and *C. sorokiniana* UTEX 1602 (bottom) when grown at pH 5.7 versus growth at pH 7.6. Values are plotted on a log scale for each axis. Only genes that registered a nonzero value for all 6 samples tested in this analysis were included in the plots.

Global Transcriptional Changes

The global transcriptional change for both *M. conductrix* and *C. sorokiniana* were compared when grown at pH 5.7 versus 7.6 (Figure 1.6). The number of significantly differentially transcribed genes (either more than a two-fold increase or decrease with a p-value < 0.05) is similar for both strains, 155 genes total for *M. conductrix* (41 up and 114 down) and 139 for *C. sorokiniana* (103 up and 36 down). However, if non-significant highly differential genes are included in the comparison between the strains, the global transcriptional response to a slightly acidic environment is more specific for *M. conductrix* when compared to *C. sorokiniana*. This can be seen in the volcano plot as there is a stronger clustering of genes about the zero value on the X-axis (fold change) in *M. conductrix*, while the dispersion of genes is greater for *C. sorokiniana.* This might be expected if the strain had evolved to function within an environment where the pH was slightly acidic or even fluctuating between two alternating pH ranges, if the host species modulates the pH within the perialgal space to trigger extracellular sugar release.

DISCUSSION

Potential Mechanisms of Extracellular Maltose Transport

In lieu of any direct evidence of which gene or protein is responsible for transporting maltose outside of the cell, we used examples of sugar transport in other organisms as indirect evidence to propose possibilities for the mechanism utilized by *M. conductrix*.

The simplest method to achieve maltose export without the sugar just leaking out of the cell would be the simple expression of a maltose specific uniporter that merely allows the maltose which has accumulated in the cytosol to travel out of the cell through a mechanism of facilitated diffusion based on a differential sugar gradient (Figure 1.7A). An example of this type of passive uniporter is MEX1 (Maltose Excess 1) which localizes in the chloroplast membrane, allowing maltose to move from the chloroplast to the cytosol (Niittylä *et al.*, 2004). A homolog of MEX1 was found in both of our newly sequenced genomes (Table 1.3), but only one copy of the gene is present and there is no

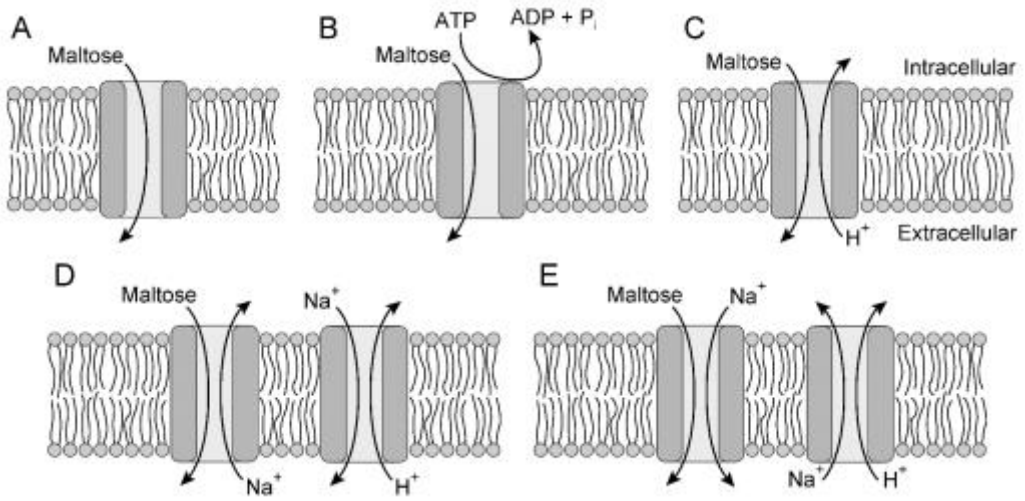


Figure 1.7. Potential mechanisms for the transport of maltose outside of the cell in *Micractinium conductrix* SAG 241.80. (A) Uniporter that allows facilitated diffusion from an elevated concentration of maltose inside the cell. (B) Uniporter powered by active transport using stored chemical energy from within the cell. (C) Antiporter utilizing the proton gradient that is due to the pH differential between the extracellular and intracellular spaces. (D) and (E) Coupled cotransport systems that would use the proton gradient to create another ion gradient (here Na^+) that would then be used to drive a sugar cotransporter to move maltose out of the cell.

evidence that the transporter localizes in both the chloroplast membrane and cell membrane. Based on the phenotype confirmed here expression of this protein would be increased by an external stimulus sensing the lower pH of the environment, ultimately resulting in the elevated levels of maltose released. This differential expression should then be evident in the RNA-Seq transcriptional studies. While our analysis of theRNA-Seq results did identify a pool of genes that were differentially expressed when grown at the lower pH 5.7, none of these genes were annotated or specifically attributed to a known sugar uniporter family.

The potential transport protein could still be a uniporter without being differentially expressed if it utilized stored chemical energy, like the dephosphorylation of ATP (Figure 1.7B). The protein could be constitutively expressed but only actively transport maltose based on a response to the environmental conditions, likely through an allosteric mechanism or signal cascade. Previous studies have claimed that maltose or glucose release in certain symbiotic species of *Chlorella* is an active process (Fischer *et al.*, 1989, Matzke *et al.*, 1990), but while active maltose transporters are known in nature they are generally associated with the uptake of maltose in bacteria (Oldham & Chen, 2011). This includes the MalFGK complex characterized in *Escherichia coli* (Lippincott & Traxler, 1997). We did look for homologs of the MalFGK complex in our genomes (Table 1.4) but there were no BLAST results for MalF or MalG. The ABC 'ATP-Binding Cassette', MalK, did yield one hit in *M. conductrix* (g5380) and three hits in *C. sorokiniana* (g4174, g6121 and g8254) but none of the BLAST results were very well conserved as all of the E-values were higher than $1e^{-5}$. We compared these genes with MalK using Multalin and they all contained conserved domains that are found in the ABC transporter family, including the family's unique signature motif, Walker-A sites, Walker-B sites, and D-loops (Schneider & Hunke, 1998).

It is also possible that the potential transport protein isn't just signaled by the low pH but actually utilize the proton gradient as a driver of maltose transport, a maltose/H^+ antiporter system for example (Figure 1.7C). Precedence for sugar antiporters comes from a recently identified family of proteins called tonoplast sugar transporters (TSTs) which are responsible for vacuolar sugar uptake in sugar beet taproots and couple the

import of sucrose into the vacuole to the export of protons (Chen *et al.*, 2015, Jung *et al.*, 2015). The benefit of this system is that it provides a direct mechanism for the transport of sugar against a gradient. It requires no extra signaling, regulation or other proteins to function. In the BLAST results for TST homologs in our algae genomes, multiple genes were identified as possibly related to TSTs (Table 1.4), including the only gene in the table that had a significant transcription level higher at the lower pH (2 fold higher). The annotations pegged these homologs as hexose or inositol transporters.

Other potential coupled sugar transporter complexes could utilize the proton gradient as well. Sugar symporters are common in higher plants, including the sodium-glucose symporters (SGLTs), sucrose/H^+ cotransporters (STP)s and sugar transporters (SUTs) (Chen *et al.*, 2015) of which we searched for homologs and did not find any in either of our algal genomes. A cotransporter could be expressed that uses the proton gradient of the low pH environment to create a potential gradient of a non-H^+ ion, like Na^+, across the membrane. Along with this sodium hydrogen cotransporter, the cell would also need to have a cotransporter powered by the sodium gradient to export the maltose (Figure 1.7D, 1.7E). One gene that did increase its level of transcription in *M. conductrix* when grown in lower pH conditions was g7595 which is annotated as a sodium hydrogen exchanger (g7595, Table 1.4), though this antiporter would still require the presence of a sugar antiporter for us to see the maltose excretion phenotype. There is also no evidence to demonstrate that this sodium hydrogen exchanger is located in the correct membrane or orientation to facilitate a coupled system. These potential coupled-cotransporters could be based on any number of secondary ions, but are notably more complex that simply utilizing a maltose/H^+ antiporter. These alternatives are highly speculative in nature and were included to illustrate some of the possibilities that may need to be explored in the attempt to identify the transporter responsible for extracellular maltose excretion in *M. conductrix*.

Chapter 2
Isolating Other Sugar Producing Algae

INTRODUCTION

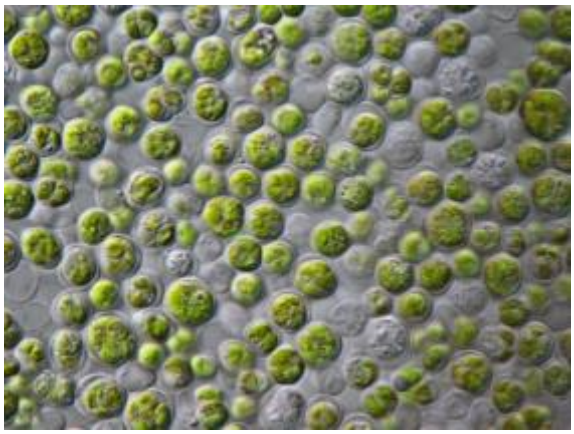


Figure 2.1. PABB004 *Scenedesmus* sp. imaged with a light microscope

In order to determine what gene or genes may be important for extracellular sugar production, it would be helpful to have additional sequenced algae genomes that have a sugar excreting phenotype for comparison to one another and to *Micractinium conductrix* SAG 241.80, which we have already sequenced and annotated. Obtaining a broader collection of sequenced strains that exhibit this phenotype would allow us to search for additional potential gene targets with a greater degree of certainty, and could help to reveal multiple genes able to accomplish extracellular sugar release. As part of this effort we decided to collect and culture additional environmental isolates of algae from *Paramecium bursaria* collected from local ponds and lakes.

In this chapter, we have described the isolation of four strains of green algae and characterized these strains in relation to their ability to produce extracellular sugars. One strain that we isolated, PABB004 *Scenedesmus* sp. (Figure 2.1), produced maltose and glucose levels five times what was seen in *M. conductrix* at pH 7.0. This green algae strain has been sequenced with both PacBio and Illumina and is currently being assembled and annotated.

MATERIALS AND METHODS

Environmental Sampling and Strain Isolation

Water samples were collected (1.0 L) from five freshwater ponds (named pond A through E) in and around Minneapolis and St. Paul, MN, USA. Samples were first run through a sterile funnel containing a cotton ball plug to remove any larger debris and collected in an autoclaved glass bottle. Micronutrients were added along with several grains of autoclaved rice. After several weeks of growth at room temperature under standard laboratory lighting, 5 mLs of solution from the culture was transferred to 250 mL of Paramecium Growth Medium. Cultures were transferred to fresh medium every 3 months to maintain the cultures. *Paramecium* were further purified by inserting a small plug of compact cotton into an autoclaved glass pipette that had been trimmed to have a 1 inch spout and 2 inch large barrel. Culture containing *Paramecium* was added to the larger barrel side of the cotton plug and the modified pipettes were stored horizontally on a light table for up to a week to allow the Paramecium to swim through the cotton plug and separate themselves from non-motile algae within the culture. The motile *Paramecium* were then collected from the spout side of the glass pipette using a 10 μL sterile plastic pipette tip, and were spotted onto an agar plate containing Freshwater SAG Growth Medium. Plates were visually inspected to determine spots (approximately 0.5 μL each) that contained *Paramecium* cells, and the *Paramecium* were allowed to lyse and spill their algal contents. Algal colonies that developed from the spot where a *Paramecium* cell had ruptured were assumed to be derived from the *Paramecium* and were transferred with a sterile toothpick on to a new plate of Freshwater SAG growth medium and streaked to isolate. The algae continued to be transferred to new plates successively until visually clean of contamination. All algal growths on Freshwater SAG Plates were grown on fluorescent or LED light beds at room temperature.

Paramecium Growth Media

Paramecium were grown in liquid Paramecium growth medium (0.24 mM $Ca(NO_3)2{\cdot}4H_2O$, 0.14 mM KNO_3, 0.06 mM $MgSO{\cdot}7H_2O$, 0.1mM KH_2PO_4, 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer, .05 mM soy lecithin, and 1 wheat seed grain.

Freshwater SAG Growth Media and Plates

Algae samples were maintained on solid Freshwater SAG growth medium plates (1.4 mM K_2HPO_4, 0.3 mM $MgSO_4{\cdot}7H_2O$, 0.17 mM $CaCl_2{\cdot}2H_2O$, 0.43 mM NaCl, 1.5 mM Na_2SO_4, 27 µM $(NH_4)_5Fe(C_6H_4O_7)_2$, 7.1 mM $NaNO_3$, 12 mM $Na_2S_2O_3{\cdot}5H_2O$, 100 µL vitamin solution [50 mg thiamine HCl, 1 mg biotin and 1 mg cyanocobalamin in 10 mL dH_2O], 2 mL BME vitamin solution [P/N B6891, Sigma Aldrich, St. Louis, MO], 1 mL trace element solution [1 g boric acid, 1 g sodium EDTA, 200 mg $MnCl_2{\cdot}4H_2O$, 20 mg $ZnCl_2$, 15 mg $CuCl_2{\cdot}2H_2O$, 15 mg $Na_2MoO_4.2H_2O$, 15 mg $CoCl_2{\cdot}6H_2O$ and 10 mg KBr in 1 L dH_2O] and 12 g of plant grade agarose [P/N A7921, Sigma Aldrich], pH 7.6) at 22 °C. Modified freshwater SAG medium was used as the liquid media for experiments (1.4 mM K_2HPO_4, 0.3 mM $MgSO_4{\cdot}7H_2O$, 0.17 mM $CaCl_2{\cdot}2H_2O$, 0.43 mM NaCl, 1.5 mM Na_2SO_4, 30 µM $(NH_4)_5Fe(C_6H_4O_7)_2$, 3.5 mM $NaNO_3$, 4 mM $Na_2S_2O_3{\cdot}5H_2O$, 50 µL polypropylene glycol, 200 µL vitamin solution, 2 mL BME vitamin solution and 3 mL trace element solution into 1.4 L dH_2O, pH 7.6). For low pH experiments (pH 5.7) 10 mM MES (2-(*N*-morpholino)ethansulfonic acid) was added as buffer. For neutral pH experiments (pH 7.0), 10 mM MOPS (3-(N-morpholino) propanesulfonic acid) was added as buffer.

Algal Growths in Tubular Photobioreactors

Algal cultures were grown (N=3) in glass tubular photobioreactors with an internal volume of 1.5 L (5 cm inner diameter by 80 cm length) containing a long capillary tube (1 mm inner diameter by 75 cm length) which was used to supply 0.2% CO_2 enriched air at a constant flow rate of 0.3 L gas per L of solution per minute. The aeration served as a source of CO_2, provided oxygen exchange and was the sole source of mixing to the culture during the experiments. The algal cells were grown on agar plates at room temperature on LED light beds. Once a sufficient mass of cells had grown, the cells

were scraped with a sterile loop and inoculated into liquid media. Cultures were grown at 22 °C with a 14-hour light and 10-hour dark cycle, to mimic a natural light cycle. 2 mL samples were drawn from each tube daily for 21 days. The sample was measured for optical density (OD) at wavelength (λ) 750 nm, then centrifuged (2 minutes at 20000 g) and the supernatant was saved and frozen (-20 °C) for future analysis. Algae were grown at a standard pH (7.0) and a low pH (5.7).

Sugar Concentration Analysis

Enzymatic assays were used to quantify the concentration of both glucose and maltose excreted by the algae into the media. Glucose was determined by putting 200 µL of the thawed supernatant sample into a plastic cuvette and mixing in 1 mL of the glucose assay reagent from the sucrose assay kit (Sigma Aldrich, St. Louis, MO) without adding invertase. After a 15-minute incubation period at room temperature, the OD at λ 340 nm was measured and compared to a standard curve to determine the concentration of glucose that was in the sample. The supernatant would be diluted if the OD_{340} reading was outside the range of the standard curve. Maltose was determined by first mixing the 200 µL of supernatant with 100 mM potassium phosphate buffer pH 6.6 with 100 mM KCl and α-glucosidase (Sigma Chemical Company P/N G5003) and incubating at 37 °C for 30 minutes. The 1 mL of glucose assay reagent is then added without invertase, the sample is incubated for 15 minutes at room temperature and the OD_{340} is measured.

Genomic DNA Isolation

Scenedesmus sp. PABB004 cells were scraped with a sterile loop from agar plates and isolated using the ZR Fungal/Bacterial DNA Microprep kit as directed by the manufacturer (Zymo Research, Irvine, CA). Following isolation, the DNA quantity and quality was evaluated using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA).

RNA Isolation

Scenedesmus sp. PABB004 cells were collected in triplicate at a time in which little to no extracellular sugars were being produced (Day 5) and a time where extracellular sugars were being produced (Day 13) and frozen in liquid nitrogen for total RNA preparation. Briefly, cells were harvested by centrifugation at 14000 g for 1 minute and cell pellets were frozen in liquid nitrogen then stored at -80˚C. Frozen pellets were suspended in 1 mL Trizol and transferred to a bashing beads tube (ZR Fungal/Bacterial DNA Microprep kit, Zymo Research, Irvine, CA). The solution was vortexed with the beads for 2 minutes, then incubated at room temperature for 5 minutes. Next, 200 µL of chloroform was added to the mixture, vortexed well and centrifuged (12000 g for 3 minutes). The upper layer was carefully collected in an RNase-free tube and mixed with an equivalent volume of 95% ethanol (Sigma) and vortexed well. The solution was further purified using the Direct-zol RNA kit (Zymo Research, Irvine, CA) as directed by the manufacturer. The total RNA was collected and treated with RNase free DNase I (Zymo Research, Irvine, CA) for 10 minutes and then purified again using the Direct-zol RNA kit. Total RNA quantity and quality were measured using the NanoDrop 2000 spectrophotometer.

Genome Sequencing

Genomic DNA of *Scenedesmus* sp. PABB004 was sequenced using the Illumina Hi-Seq 2500 (Illumina, San Diego, CA) at the University of Minnesota Genomics Center using the standard protocol for paired-end reads. *Scenedesmus* sp. PABB004 was also sequenced using Pacific Biosciences (PacBio) Single Molecule Real Time (SMRT) technology at the Rochester Mayo Medical Genome Facility. 5 µg of raw genomic DNA was analyzed for size, quality and concentration by the Qubit dsDNA BR Assay (Invitrogen, Carlsbad, CA) and Fragment Analysis (Advance Analytical Technologies, Inc., Alkeny, IA) using the High Sensitivity Large Fragment Analysis kit. SMRTbell template libraries were generated using the SMRTbell Template Prep Kit 1.0. All purification and concentration steps were performed using AMPure PB beads (Pacific Biosciences, Menlo Park, CA). BluePippin Size Selection (Sage Science, Beverly, MA) was performed to optimize large insert libraries for improved long-reads. The annealing and binding reactions of BluePippin Size-Selected SMRTbell templates were carried out

using the DNA/Polymerase Binding Kit P6 (Pacific Biosciences, Menlo Park, CA). The Annealing and Binding Calculator (version 2.3.1.1) provided by Pacific Biosciences was used to calculate the concentration of bound complex to be loaded onto the sample plate and to determine which MagBead Kit (Pacific Biosciences, Menlo Park, CA) loading steps and recipes were ideal. DNA samples were run on the PacBio RSII SMRT DNA Sequencer (Pacific Biosciences, Menlo Park, CA) using the DNA Sequencing Reagent 4.0 kit and SMRT Cell v3 cells. Instrument parameters were setup as follows: MagBead collection protocol, 20,000 bp insert size, stage start and 240-minute movie collection time.

RNA Sequencing

Poly(A)$^+$ tag based mRNA isolation and RNA sequencing were performed at the University of Minnesota Genomics Center. Poly(A)$^+$ RNA was isolated using oligo(dT) magnetic beads and isolated mRNA was fragmented into small pieces at elevated temperature. The fragmented mRNA was used as template for cDNA synthesis using random primers. Further, 5' and 3' adapters were ligated to the mRNA and the adapter-ligated RNA was reverse-transcribed and amplified using PCR. The final PCR products were size fractioned and used for cDNA sequencing library construction. The cDNA sequencing libraries were sequenced using the HiSeq 2500 Illumina sequencing technique.

RESULTS AND DISCUSSION

Paramecium Derived Algae Strains Isolated

Four strains of algae were successfully isolated from the *Paramecium* collected from the five ponds sampled. The 18s rRNA sequence of each alga was determined by performing PCR with degenerative primers, and the PCR products were purified and sequenced by Sanger sequencing to yield 800-1000 bp of sequence which were used to identify the strains. The closest match for the genus based on each sequence is shown in

Table 2.1. List of environmental algae strains isolated

Strain	Isolated From	Algae	Sugar Production?
PABB001	Pond B	*Micractinium* sp.	No
PABB002	Pond D	*Coelastrum* sp.	Yes, at pH 5.7
PABB003	Pond E, Colony 1	*Scenedesmus* sp.	No
PABB004	Pond E, Colony 2	*Scenedesmus* sp.	Yes, at pH 7.0

Table 2.1. Preliminary growth analysis and sugar concentration assays (data not shown) were performed to determine if these isolated strains produced any traceable amount of extracellular sugars. Two of the strains were found to produce extracellular sugars: PABB002, identified as *Coelastrum* sp. at pH 5.7 and PABB004, identified as *Scenedesmus* sp. at pH 7.0. It should be noted that PABB004 does not exhibit the general phenotypical markers that are associated with *Scenedesmus* algae and that this classification could change. Unfortunately, before we were able to perform more stringent experiments to quantify extracellular sugar production by PABB002, the laboratory stock of the strain died and we were unable to resuscitate it or revive it from frozen or freeze-dried preparations. Attempts to re-isolate this strain are underway.

Sugar Analysis

We grew the environmental isolate PABB004 in triplicate for 21 days to quantify the type and amount of extracellular sugars it produced (Figure 2.2). PABB004 produces primarily glucose and a lesser amount of maltose, which is in stark contrast to what is found to be produced by *Micractinium conductrix* SAG 241.80. Another significant difference seen between PABB004 and *M. conductrix* is in the amount of sugar that was produced. The highest concentration of sugars *M. conductrix* excreted was 700 μM maltose and 50 μM glucose (1.45 μM glucose equivalent, as maltose is comprised of two molecules of glucose), whereas PABB004 produced 1.1 mM maltose and 5 mM glucose (7.2 mM glucose equivalent), a level which is five-fold higher. Sugar excretion occurred more slowly initially in PABB004, where negligible amounts of extracellular sugar were

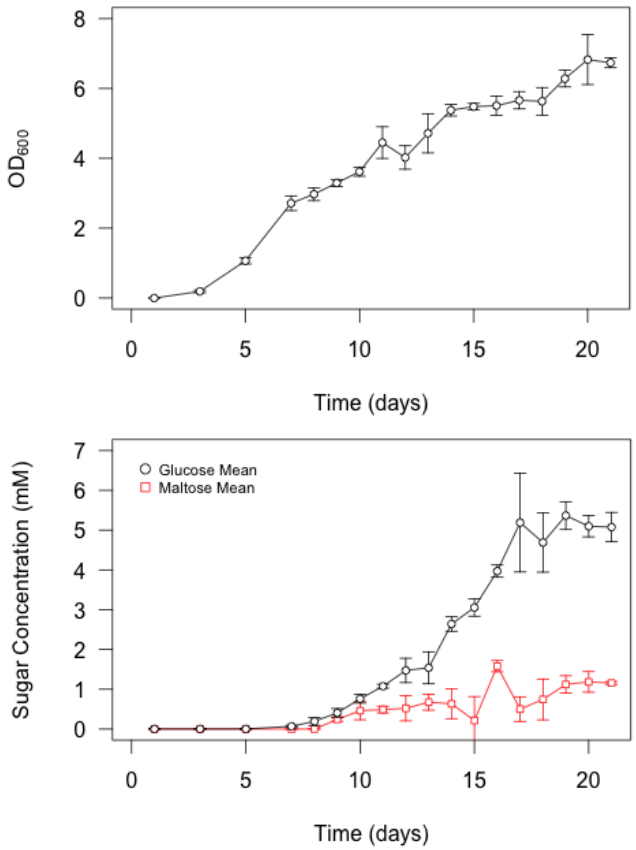


Figure 2.2. Sugar Accumulation by PABB004 *Scenedesmus* sp. Upper panel shows the growth of PABB004 *Scenedesmus* sp. while the lower panel shows the amount of glucose (●) and maltose (□) the green algae produced (N=3).

---

detected until day 8 compared to *M. conductrix* which had extracellular sugar levels gradually increase until leveling out at day 8. Perhaps the most important distinction found for the PABB004 strain compared to other reported sugar-secreting algae is related to the pH at which sugars are released. We were unsuccessful in culturing PABB004 at lower pH values such as pH 5.7, which is used to yield sugars in *M. conductrix*. However, PABB004 was able to grow at pH 7.0, and so the initial evaluations of extracellular sugar production were performed at this pH. The identification of a strain of algae that produces high levels of sugar at a more physiologically neutral pH could have broader application in developing successful co-cultures or artificial symbiotic relationships. The fact that PABB004 released sugars at pH 7.0 could also indicate the

possibility of a different mechanism for sugar release than what was discussed for *M. conductrix* in the previous chapter.

FUTURE DIRECTIONS

Genome Sequencing, Assembly and Annotation

The genome of PABB004 *Scenedesmus* sp. was sequenced with both PacBio and Illumina, in a similar manner to what was done previously with *M. conductrix* SAG 241.80 and *C. sorokiniana* UTEX 1602. Assembly of the *de novo* genome is currently underway. A preliminary RNA-Seq experiment was also performed, with samples taken before sugar production was seen as well as during sugar production in an attempt to highlight genes that may play a role in the extracellular sugar production phenotype. The RNA-Seq data will also be used to help the annotation of the genome which will be performed using funannotate[6].

Continue Isolation and Sequencing

A current effort is underway to continue isolating environmental strains of algae as has been discussed here. These strains will also be investigated for extracellular sugar production. Any strains that do produce extracellular sugars (and have not been previously sequenced) will be sequenced and annotated to obtain a *de novo* complete genome. We will use these new genomes to continue our search for the gene or genes responsible for releasing extracellular sugars as discussed in the previous chapter.

[6] https://github.com/nextgenusfs/funannotate/wiki

Chapter 3
The Design of a Low Cost Turbidostat

INTRODUCTION

A turbidostat is a continuous microbiological reactor in which the density or turbidity of the culture growing inside is held constant. This is achieved by introducing new media to dilute the culture if it becomes too dense. The growth rate of the culture can be inferred by how much new media is being introduced to the reactor. The faster that new media is being pumped into the reactor, the higher the rate of growth for the organism or organisms inside. The operation of the turbidostat is described by a simple logic diagram (Figure 3.1). The turbidity of the culture is routinely measured, and if it is above the set point defined by the user then media is added to dilute the culture. If the turbidity of the culture is below the set point, then no action is taken.

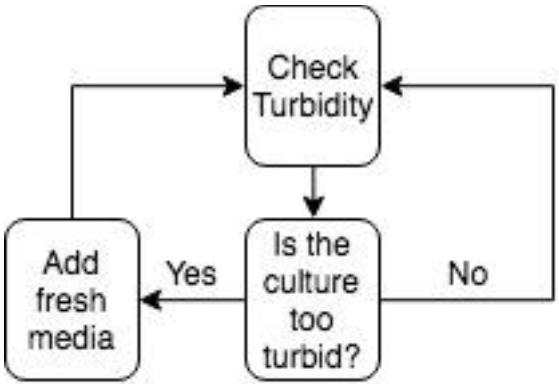


Figure 3.1. Flow chart representing the logic that defines a turbidostat.

Turbidostats are very useful for running evolutionary experiments (Blumwald & Tel-Or, 1984, Toprak *et al.*, 2011 Avrahami-Moyal *et al.*, 2012), which are often run in batch culture due to cost and relative simplicity. However, performing batch culture evolutionary experiments can add unwanted selective pressure to the population as the physiology changes throughout the various stages of growth during the experiment (Hengge-Aronia, 1993, Brauer *et al.*, 2008, Klumpp *et al.*, 2009). As turbidostats add

fresh media and maintain the culture at a constant culture density, the cells are theoretically kept in a consistent environmental condition. Selective pressures, such as a toxin, can be added to the incoming media in incrementally higher doses to induce evolution. However, commercial turbidostats can be costly and even other published "low-cost" solutions were fairly expensive ($1400-$2000, Takahashi *et al.*, 2015, Matteau *et al.*, 2015) so we decided to design and build our own simple turbidostat.

As a proof of concept, we wanted to perform a directed evolution experiment in the turbidostat we designed and assembled to make sure the reactor functioned as planned. We chose to run an experiment that would force the bacterium *Azotobacter vinelandii* to utilize the toxic chemical hydrazine (N_2H_4) as its sole source of nitrogen. *A. vinelandii* is a model organism for the study of nitrogen fixation by an aerobic Gram-negative bacterium (Setubal *et al.*, 2009). This model organism contains three nitrogenase enzymes which convert atmospheric dinitrogen gas into ammonia: the molybdenum-iron (MoFe), the iron only (Fe) and the vanadium (V) nitrogenases (Barney *et al.*, 2006, Hoffman *et al.*, 2014). The nitrogenase enzyme is one of the strongest reducing enzymes found in nature and is able to act upon a number of other substrates besides dinitrogen gas, including acetylene, cyanide, azide, carbon monoxide, carbon dioxide and hydrazine (Burgess & Lowe, 1996). Hydrazine, however, is a very poor substrate for nitrogenase with a K_m that is highly dependent on pH (Davis, 1980). The decision was made to focus solely on evolving the MoFe nitrogenase because it has been shown that specific, single amino acid mutations to the NifD subunit of the MoFe protein have been found to dramatically increase its ability to reduce hydrazine *in vitro* (Barney *et al.*, 2004). The fact that only one mutation is necessary to see a drastic change in the kinetics of the MoFe protein makes it a good target for a proof of concept evolution experiment as theoretically less time is needed to obtain an evolved strain. The *A. vinelandii* strain AZBB108 was used in the evolution experiment as this strain only contains the MoFe nitrogenase because the vanadium and iron only nitrogenases have been knocked out (*ΔvnfHDGK ΔanfHDGK*).

AZBB108 was grown under conditions where dinitrogen was not available (i.e. under an Ar/O_2 atmosphere) and hydrazine was present as the only source of nitrogen in

our turbidostat in an attempt to evolve the MoFe nitrogenase to more efficiently utilize hydrazine. If a strain did evolve to utilize hydrazine as its sole nitrogen source, we could then investigate if the mutations that led to this efficiency were the same mutations as previously reported, or were novel mutations. This would also require *A. vinelandii* to be able to grow in the presence of hydrazine, which in preliminary studies was found to be toxic at 0.5 mM when grown on plates.

The goal of this endeavor was to create a cost-effective ($430 not including a stir bar and 1L and 2L bottles that are common in labs), user-friendly turbidostat that can easily be assembled by any laboratory or classroom that has the necessary components and access to a 3D printer. We performed a proof of concept directed evolution experiment that in the end resulted in a strain that could grow in the presence of hydrazine but could not grow utilizing hydrazine as its sole nitrogen source. Importantly, this experiment proved that the turbidostat constructed for this application does work as designed.

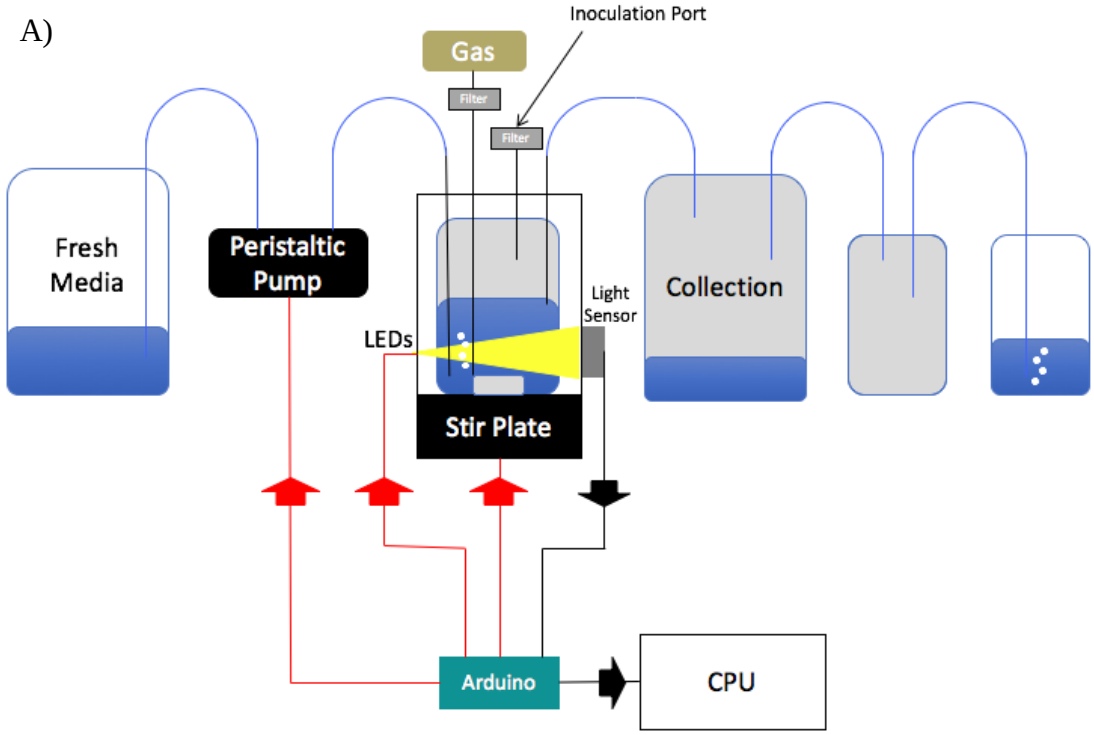
MATERIALS AND METHODS

Turbidostat Design

The turbidostat is comprised of multiple parts (Figure 3.2, Table 3.1). The reactor itself is a 500-mL square bottle (Pyrex, Rosemont, IL) with a size 7 rubber stopper (Karter Scientific, Lake Charles, LA) that make it possible to isolate the atmosphere within the reactor from the external atmosphere so that it can be controlled and manipulated. Three holes are drilled into the rubber stopper large enough to slide metal capillaries with fittings (Swagelok, Chaska, MN) through, but tight enough that gas from inside the reactor cannot escape. These three capillaries were used as a medium inlet, a gas inlet, and a gas/spent medium outlet. A 20G x 1.5” needle (Becton-Dickinson, Franklin Lakes, NJ) is also physically pushed through the stopper and acts as the inoculation port. Silicon tubing (Thermo Fisher, Waltham, MA) connects the fresh media bottle to the reactor bottle, to the collection bottle, to the overflow bottle, and to the final bottle, constituting a discrete, closed system referred to as the “bottle system”. Media is introduced from the fresh media bottle via a peristaltic pump (Boxer, Ottobeuren,

Germany) which is a positive displacement pump that only allows flow while pumping. The pump is constantly engaged with the specialized Boxer pump tubing, effectively eliminating any backflow. All of the 3D printed pieces were designed using the free online software Tinkercad[7].

A)



B)

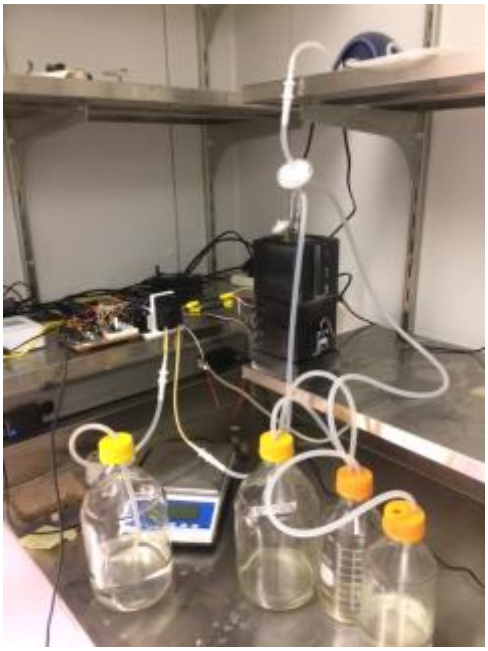


Figure 3.2. The components that comprise our turbidostat system. (A) Graphical representation. Red arrows going from the Arduino microcontroller show what components are being controlled via our code. Black arrows going to the Arduino and CPU represent the collected and saved data (lux readings). Grey, shaded in bottles represents the isolated atmosphere inside of the turbidostat. (B) Image of the turbidostat running an experiment.

[7] https://www.tinkercad.com/

Table 3.1. Parts list for turbidostat

Tubing/Bottles	Per Reactor	Part Number	Manufacturer	Purchased From
Silicone Tubing 3/16"x5/16"	12 ft.	8060-0040	Thermo F	Thermo F
Stainless Steel Tubing 1/8"x.028" Wall (Capillaries)	1 ft.	SS-T2-S-028-10	Swagelok	Swagelok
Hose Connector Adapter 1/4"x1/4"	3	SS-4-HC-A-401	Swagelok	Swagelok
Reducing Union 1'4"x1/8"	3	SS-400-6-2	Swagelok	Swagelok
2 L Pyrex Bottles	2	1395-2LCNEa	Pyrex	Amazon
1 L Pyrex Bottles	2	1395-1LCNEa	Pyrex	Amazon
Size 7 Rubber Stopper	1	216R4	Karter Scientific	Amazon
Air Pump	1	77851	Tetra	Amazon
Electronics				
Arduino Uno	1	A000099	Arduino	Arduino
USB 2.0 Cable A-Male to B-Male	1	PC045	Amazon	Amazon
Heat Shrink Tubing	8	60312WDF	Vktech	Amazon
18-Gauge Red Wire	2 ft.	55667423	Southwire	Amazon
18-Gauge White Wire	2 ft.	55667223	Coleman Cable	Amazon
Breadboard Jumper Wires	26	EA-BJW-A-=130	Sim & Nat	Amazon
Mini Breadboard	1	MK-003	Qunqi	Amazon
5mm Ultra Bright LED Lights	3	H&PC-54378	Blucell	Amazon
5 mm LED Light Holder Mount	3	a14051900ux0209	uxcell	Amazon
Light Sensor	1	P/N 1143	Phidgets	Phidgets
100 Ω Resistors (for LEDs)	3	B0185FCR66	Xicon	Arduino
10 kΩ Resistor (for Pushbutton Switch)	1	B00CVZ4DKU	Xicon	Arduino
Pushbutton Switch	1	CSA0150	Mouser	Arduino
Media Pump				
9K Peristaltic Pump w/ 24V Stepper Motor	1	9600.001	Boxer	Clarksol
UIM24032B Control Board	1	6900.001	UI Robot	Clarksol
Pump Tubing	1 ft.	9000.537	Clarksol	Clarksol
Tube Clamps	2	9000.603	Clarksol	Clarksol
24V AC/DC Power Cord	1	3206-24V	LEDwholesalers	Amazon
DC Power Connector	1	DC-Power-Connector-3071	Conwork	Amazon
10 kΩ Trimmer Potentiometer	1	CA9	Kennon	Arduino
Stir Plate				
40mmx10mm 5V DC Computer Fan	1	L-FENG-UK	uxcell	Amazon
BC547B Triode Transistor	1	B072NC6X4M	quickbuying	Arduino
5" Bar Magnets	2	28N029	Alinco	Amazon

The reactor is airtight as the air inlet introduces a positive pressure into the reactor, the media inlet is controlled by the peristaltic pump which does not allow any backflow, and the inoculation port is capped after the inoculation. This means that the only means for anything exiting the reactor is the outlet. In this experiment, our outlet level was set to a point resulting in an internal reactor volume of 300 mL. The positive pressure in the headspace of the reactor forces all media and excess gas through the outlet into the collection bottle. The gas continues through the overflow bottle until it bubbles out the final bottle that is half filled with water, resulting in a closed atmosphere. For experiments requiring standard atmospheric conditions (21% oxygen and 78% nitrogen), an aquarium air pump (Tetra, Blacksburg, VA) is used to force filtered air through the reactor. For experiments requiring alternative atmospheres, mass flow controllers (Alicat Scientific, Tucson, AZ) were used to maintain specific flow rates.

To determine the density or turbidity of the culture in the reactor, the amount of light reaching the light sensor from the externally controlled LED lights (Bluecell, Ikeja, Nigeria) was measured in units of lux, which is a unit of illuminance. In order to minimize ambient light and achieve the most accurate lux readings, the reactor bottle was housed within a 3D printed enclosure[8] (Figure 3.3). The housing is comprised of the base in which the bottle sits, and a two-part top which allows for the bottle to be enclosed already completely connected to the bottle system via tubing. This allows for autoclaving of the whole bottle system as a unit. Temperature was maintained by placing the entire turbidostat apparatus in a constant-temperature room that was set at the temperature required for the culture to grow.

[8] https://www.tinkercad.com/things/l7TqhqHGrez-turbidostat-arriola-housing-base, https://www.tinkercad.com/things/iR66wIrcicR-turbidostat-arriola-housing-top, https://www.tinkercad.com/things/83WmUDneqrS-turbidostat-arriola-housing-door

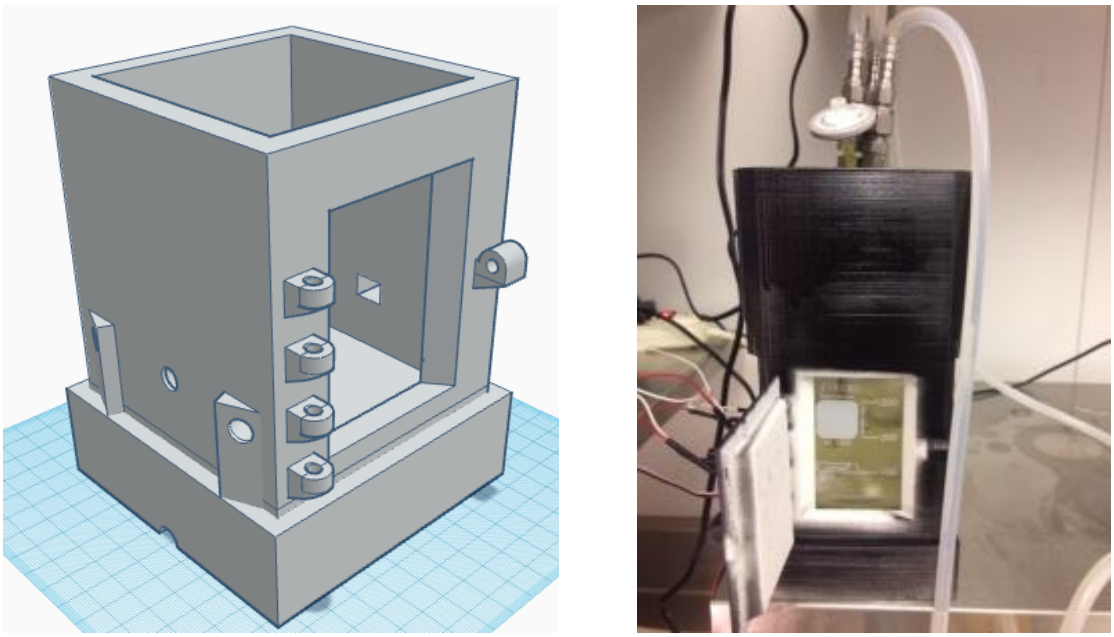


Figure 3.3. Design and print of the reactor housing. Design was done in Tinkercad[6].

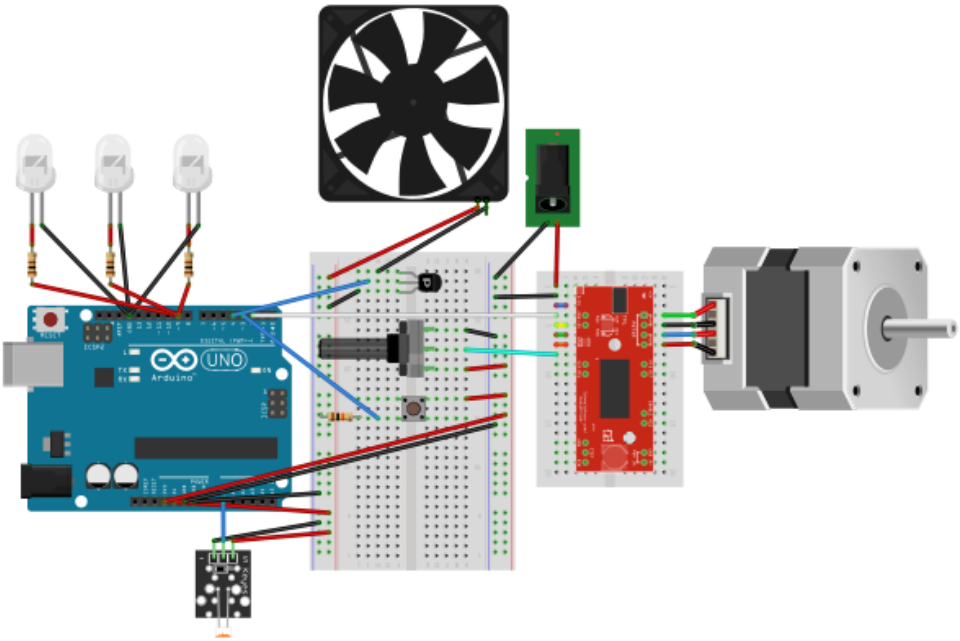
Turbidostat Components

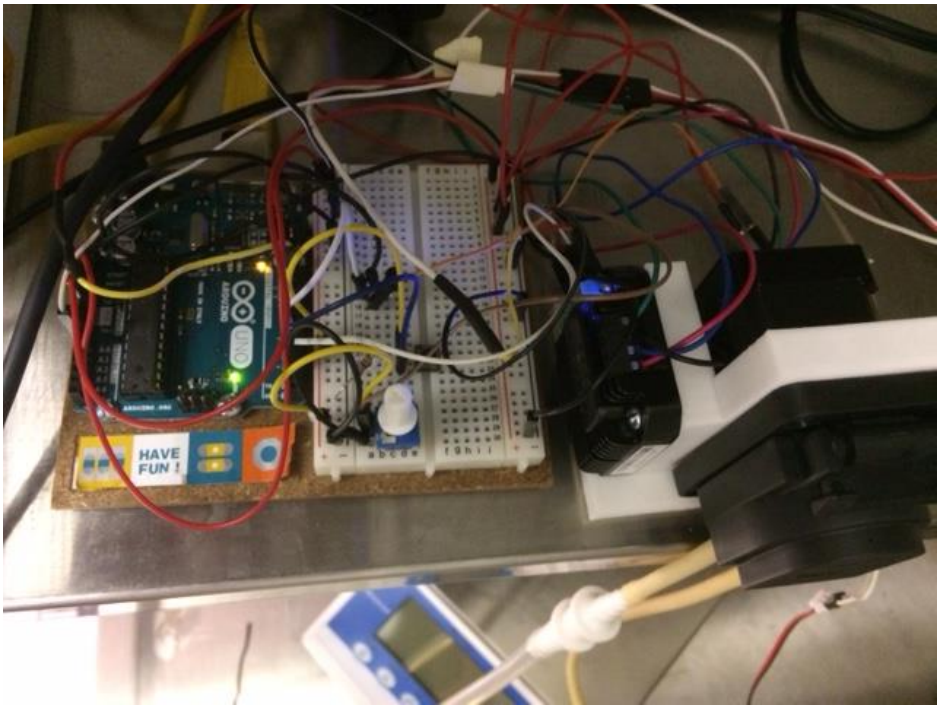
The control wiring was connected through a mini breadboard (Qunqi, Shenzhen, China) (Figure 3.4A). A peristaltic pump was used to introduce media to the turbidostat which was driven by a stepper motor as it allowed for an accurate and consistent pump rates. Communication with the motor was achieved via a stepper motor control board (UI Robot, Shanghai, China) (Figure 3.4B). The stepper motor control board and pump are screwed into a 3D printed mount[9]. The pump speed was controlled by the current trimmer on the back of the stepper motor control board which was tightened halfway. This allowed the motor to run at a speed useful for the experiment. The pump speed was also controlled by the potentiometer located on the breadboard. As there is no digital control of the pump speed, each new pump must be calibrated to determine the pump speed of that specific pump set up which must be entered into the Arduino code. The pump is powered externally (LEDwholesalers, Hayward, CA) as the 24V required to power the pump is too great to be supplied through the Arduino board.

[9] https://www.tinkercad.com/things/h1A4aMbC250-turbidostat-arriola-peristaltic-pump-with-stepper-motor-and

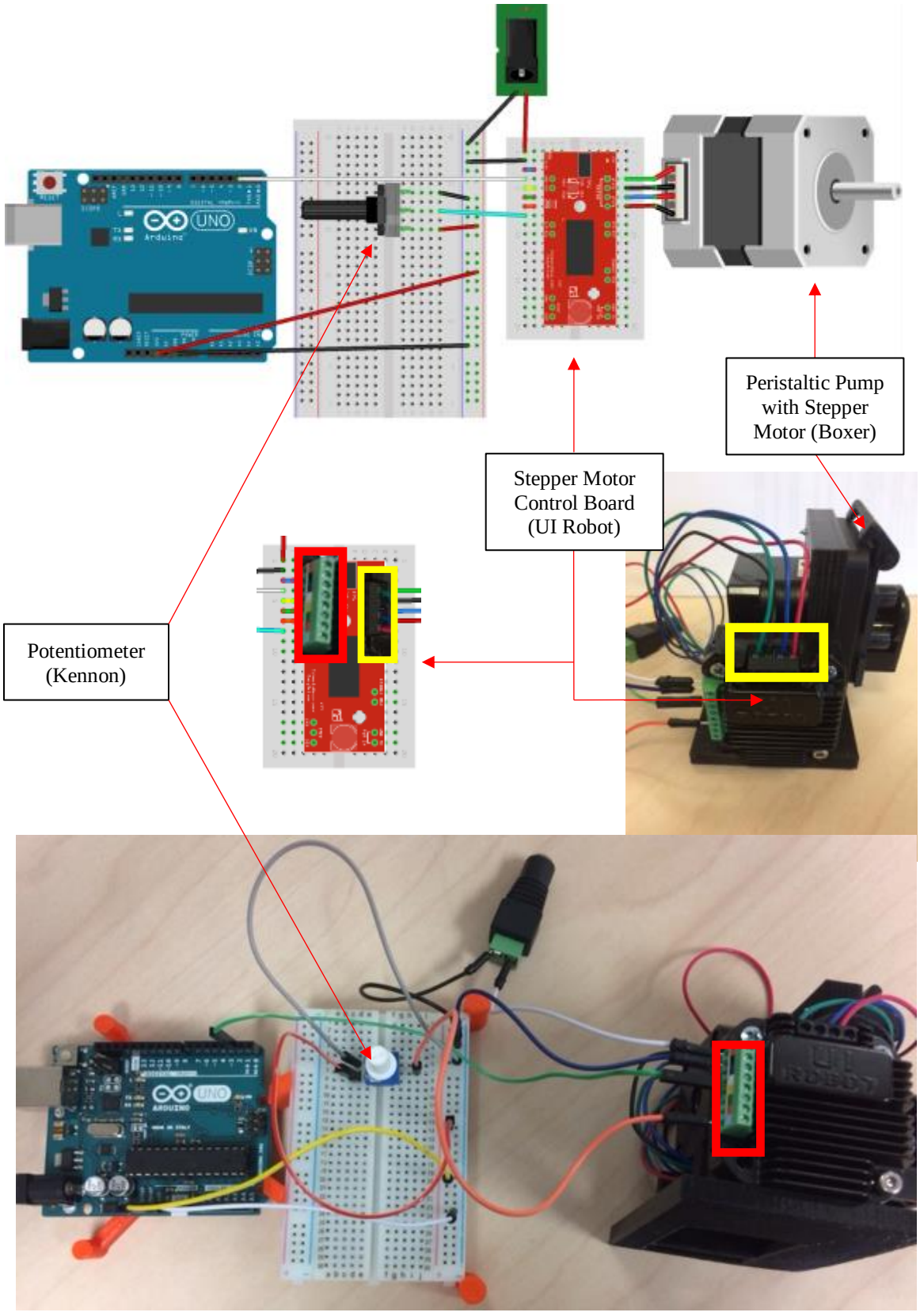
Figure 3.4. Charts and images of the components and circuits. (A) All of the components and corresponding circuits used to control the turbidostat, chart and image both provided to show the ideal and the actual setup. (B) The peristaltic pump with stepper motor and stepper motor control board, including the potentiometer that controls the speed of the pump. Note that the stepper motor control board in the wiring chart is only representative of the set up used in this system. The colors of the wires correspond to the Boxer stepper motor used in the turbidostat (see yellow and red boxes). (C) The 3 LED lights and the light sensor utilized to measure the lux. (D) The stir plate (represented here by the computer fan) and the button switch which signals to the Arduino that a new fresh media bottle has been added. Charts made in Fritzing

A)

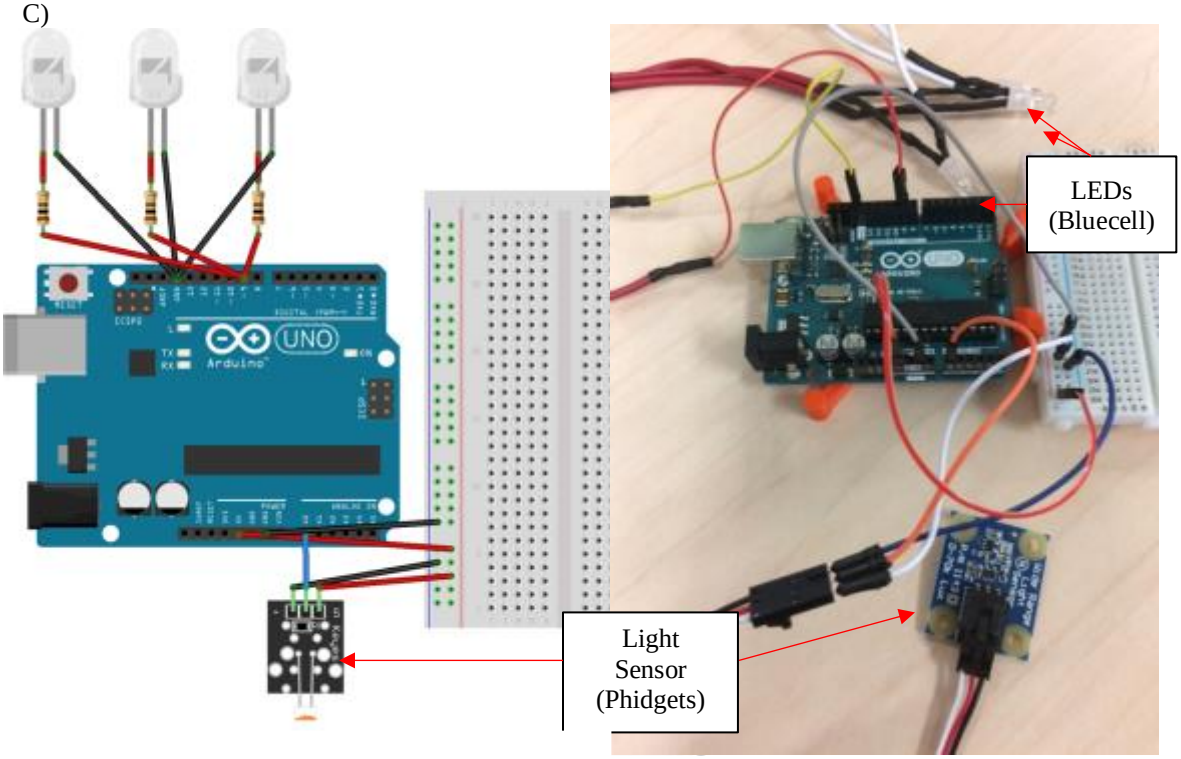




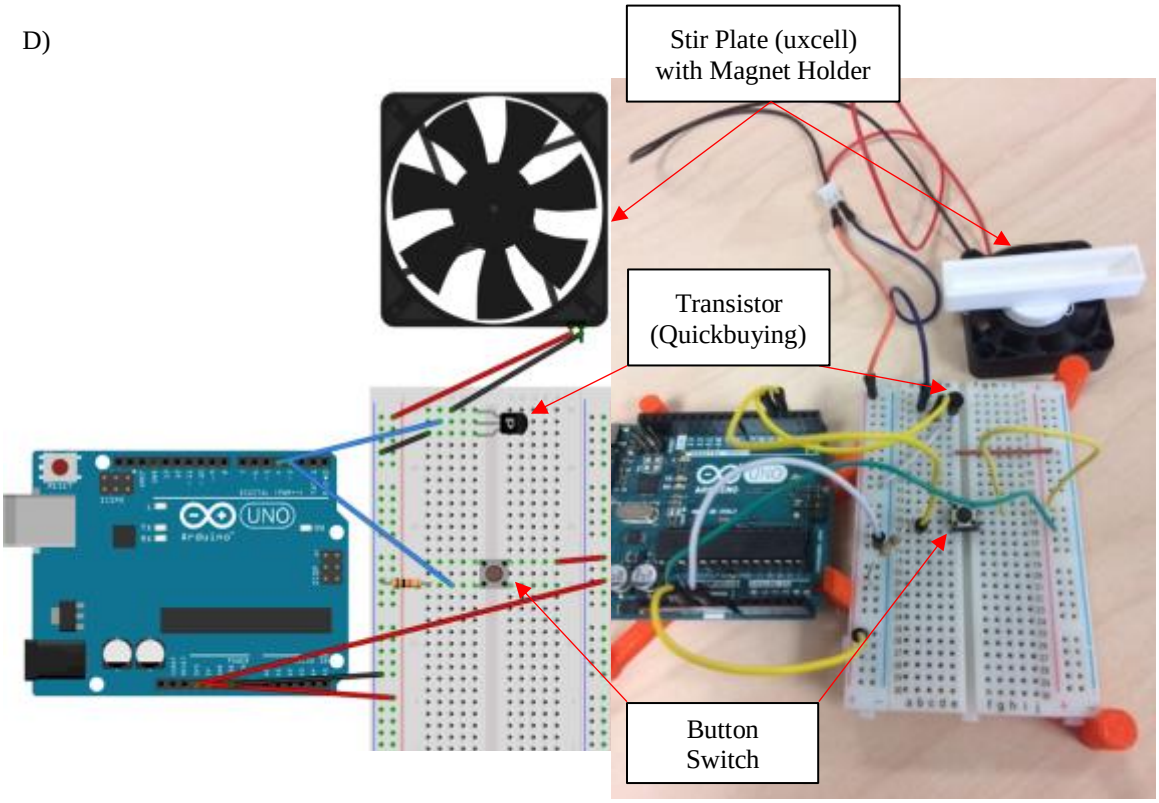
B)


Peristaltic Pump with Stepper Motor (Boxer)
Stepper Motor Control Board (UI Robot)
Potentiometer (Kennon)

C)


LEDs
(Bluecell)
Light
Sensor
(Phidgets)

D)


Stir Plate (uxcell)
with Magnet Holder
Transistor
(Quickbuying)
Button
Switch

The three LED lights (Figure 3.4C) were attached to the housing via LED holder mounts (uxcell, Hong Kong) were flashed at one second intervals three times and the light reaching the opposite side of the reactor was read by the light sensor (Phidgets, Calgary, Canada) to determine the lux. A stir plate was created by gluing a 3D printed mount[10] that holds a bar magnet (Alinco, Osaka, Japan) to a 40mM x 40mM x 10mM 5V computer fan (uxcell, Hong Kong) which is turned on and off by the Arduino via a transistor (Arduino, New York City, NY) and provided mixing for the culture inside the reactor (Figure 3.4D). A button switch (Mouser, Mansfield, TX) was also included to reset the available media count when a new bottle of media was added (Figure 3.4D)

Arduino Code

All of the components in the system are controlled by an Arduino Uno (Arduino, New York City, NY) microcontroller. A code written in Arduino script or language (Figure 3.5) is uploaded to the microcontroller and runs the turbidostat. The Arduino Integrated Development Environment (IDE)[11] is a free, open-source software platform. A flow chart of the code's loop (Figure 3.1) illustrates that it is a very simple underlying system, but complexities may be added to fit the user's need. Media is pumped until the volume of available fresh media falls below 100 mLs (which can be adjusted based on the bottle size and requirements of the user) in the fresh media bottle. Pumping until the fresh media bottle is completely empty can cause dry pumping, which is when the pump is active but is not transferring any media. This can lead to inaccurate data collection and inaccurate calculations of dilution rates. In our experimental setup, each bottle of fresh media contained 1 L of liquid, so for each new bottle of media added to the turbidostat, 900 mLs is actually transferred to the reactor.

Processing Code

The Arduino code serially prints the data to a computer connected via USB. However, Arduino does not have the ability to save data so another program must be

[10] https://www.tinkercad.com/things/5auF3dVAvsM-turbidostat-arriola-magnet-holder-for-stir-plate
[11] https://www.arduino.cc/en/Main/Software

Figure 3.5. The Arduino code used to run the turbidostat. Comments come after "//" and are bolded. Highlighted items should be changed for different experiments

```
//set point
//run pilot experiment to determine setPoint
int setPoint = 150;

//set pins
const int lightSensor = A0;
const int pump = 2;
const int stirPlate = 4;
const int mediaSwitch = 3;
const int led = 9;

//pump mLs per 10 secs
//must adjust with every new pump
float conv = 2.73;

//define variables
int readings[6];
float light;
float dark;
float lux;
float media = 1000;
float mediaTotal;
int ledState = LOW;
int i;
int bottle;

void setup() {
  Serial.begin(9600);
  pinMode(led, OUTPUT);
  pinMode(pump, OUTPUT);
  pinMode(stirPlate, OUTPUT);
  pinMode(mediaSwitch, INPUT);
  ledState = LOW;
  i = 1;
  bottle = 1; }

void loop() {

//take gross lux readings
if (i < 7) {
  ledState = !ledState;
  digitalWrite(led, ledState);
  delay(1000);
  readings[i] = analogRead(lightSensor);
  digitalWrite(led, LOW);
  i++; }

//average gross lux readings
else if (i = 7) {

light = ((readings[1] + readings [3] +
readings[5]) / 3);

dark = ((readings[2] + readings[4] +
readings[6]) /3);

lux = light - dark;

//print data to computer
Serial.print(media);
Serial.print(",");
Serial.print(mediaTotal);
Serial.print(",");
Serial.print(lux);
Serial.print(",");
Serial.println(bottle);
i = 1;
ledState = LOW;

//if OD is too high, pump media for 10 secs
//while stirring
if( lux < setPoint && media > 100) {
  digitalWrite(pump, HIGH);
  digitalWrite(stirPlate, HIGH);
  delay(10000);

//turn off pump and stirPlate
  digitalWrite(pump, LOW);
  digitalWrite(stirPlate, LOW);
  delay(5000);

  //media calculations
  media = (media - conv);
  mediaTotal = (mediaTotal + conv); }

//if OD is at or below set point OR remaining
// media  is less than 100 mLs, don't pump
else if(lux >= setPoint || media > 100) {
  delay(35000);
  digitalWrite(stirPlate, HIGH);
  delay(20000);
  digitalWrite(stirPlate, LOW);
  delay(5000); }

//when bottles are switched, press the button
if(digitalRead(mediaSwitch) == HIGH) {
  bottle++;
  media = 1000; } } }
```

Figure 3.6. The Processing code used to save data from the Arduino to the computer. Comments come after "//" and are bolded. Highlighted items should be changed for different experiments.

```
import processing.serial.*;
Serial myPort;
Table table;
String fileName;
float sensorVals[];

void setup() {
 //Number in brackets changes depending
 //on which turbidostat you wish to connect to
 //starting with zero going up (0,1,2, etc.)
 String portName =Serial.list()[0];
 myPort = new Serial(this, portName, 9600);
 table = new Table();
 table.addColumn("id");

 //the following adds columns for time
 table.addColumn("year");
 table.addColumn("month");
 table.addColumn("day");
 table.addColumn("hour");
 table.addColumn("minute");
 table.addColumn("second");

 //The following are dummy columns for each
 //data value. Add as many columns as you
 //have data values. Customize the names as
 //needed. Make sure they are in the same
 //order as the order that Arduino is sending
 //them!
 table.addColumn("media");
 table.addColumn("mediaTotal");
 table.addColumn("lux");
 table.addColumn("bottle");
}

void serialEvent(Serial myPort) {
  String val;
  val = myPort.readStringUntil('\n');

//if the read value is not null begin recording
//data
 if (val !=null) {
  println(val);
  float sensorVals[] = float(split(val, ','));
  TableRow newRow = table.addRow();

  //record time stamp
  newRow.setInt("year", year());
  newRow.setInt("month", month());
  newRow.setInt("day", day());
  newRow.setInt("hour", hour());
  newRow.setInt("minute", minute());

  newRow.setInt("second", second());

  //record sensor information. Customize the
  //names so they match your sensor column
  //names.
  newRow.setFloat("media", sensorVals[0]);
  newRow.setFloat("mediaTotal",
          sensorVals[1]);
  newRow.setFloat("lux", sensorVals[2]);
  newRow.setFloat("bottle", sensorVals[3]);

  //change fileName EVERYTIME or else you
  //will save over previous data
  fileName = "exampleFile.csv";
  saveTable(table, fileName); } }

void draw() {}
```

utilized. Processing[12] is another free, open-source software that was used to specifically read what the Arduino was communicating and saving it to the computer as a comma-separated value (.csv) file that can be analyzed by a variety of other statistical programs like Excel or R. The Processing code used to save the turbidostat data (Figure 3.6) was

[12] https://processing.org/

adapted from a code found online[13]. The end .csv file saves the amount of media remaining in the media vessel in mLs (media), the total media pumped during the run in mLs (mediaTotal), the lux (lux), and which number bottle is currently being pumped from (bottle). The bottle value, which simply increases by one with every new liter of media added, was included as an easy way to organize data for analysis post-experiment.

Turbidostat Set Up

The bottle system was autoclaved and the reactor bottle was placed in the 3D printed enclosure. The air pump was connected to the gas inlet to deliver aeration through a 0.2 μm filter placed inline and upstream of the capillary tube used for delivery. The pump tubing was clamped into the peristaltic pump. All media bottle caps were sealed tightly except for the fresh media bottle, which was connected to the bottle, but loose to avoid creating a vacuum in the bottle during delivery. For each new fresh media bottle added, the previous collection bottle was disconnected and replaced with the current fresh media bottle.

***Azotobacter vinelandii* AZBB108**

The *A. vinelandii* strain AZBB108 (ΔvnfHDGK ΔanfHDGK) was used in the proof of concept evolution experiment. This strain contains only the MoFe nitrogenase as the vanadium and iron only nitrogenases have been knocked out.

B Media

A. vinelandii was grown aerobically on standard Burk's (B) media (Strandberg & Wilson, 1968, Dos Santos, 2011). B medium contained per liter of deionized water: sucrose (20 g), $MgSO_4 \cdot 7H_2O$ (0.2 g), $CaCl_2 \cdot 2H_2O$ (90 mg), $FeSO_4 \cdot 7H_2O$ (5 mg), KH_2PO_4 (2 g), K_2HPO_4 (8 g), $Na_2MoO_4 \cdot H_2O$ (0.25 mg). Filter sterilized hydrazine sulfate ($N_2H_6 \cdot SO_4$) was added as needed to the media after autoclaving.

[13] http://www.hackerscapes.com/2014/11/how-to-save-data-from-arduino-to-a-csv-file-using-processing/

Inoculation of the Turbidostat

300 mL of sterile B medium was pumped into the autoclaved turbidostat and used as the starter culture medium. AZBB108 was grown on agar plates of B medium aerobically at 30º C for three days. Once grown, the cells were scraped with a sterile loop and resuspended in 1 mL of sterile B media. The inoculant was then introduced to the reactor via the inoculation port with a 3 mL syringe. A 0.2 μm filter was then used to cap the inoculation port to prevent contamination. The Arduino code was loaded onto the Arduino Uno microcontroller from the computer. The Processing code was run on the same computer to start data collection.

Running the Turbidostat

A. vinelandii was grown on 1.8 L of B media (two full bottles) without any hydrazine to both determine the dilution rate for a culture being grown without selective pressure, and to make sure that the pump rate used by the program was accurately reporting what was actually being transferred to the reactor. This initial B medium was then substituted with B medium containing 0.3 mM hydrazine was then added to the turbidostat. The level of hydrazine was then incrementally increased by 0.2 mM hydrazine with each new bottle of media until *A. vinelandii* was growing on 3 mM hydrazine (the last increased increment was 0.3 mM hydrazine from 2.7 mM hydrazine).

RESULTS AND DISCUSSION

Proof of Concept

The turbidostat was able to successfully run and log data during the directed evolution experiments performed on *A. vinelandii* AZBB108 while costing only $430, a fraction of the cost of turbidostat designs being published before the design process began. A representative example of what a typical experiment looked like is provided (Figure 3.7) after the data has been processed in R. Fresh medium was consumed and refilled as needed. The dilution rate was calculated by taking the rate at which fresh media was introduced to the reactor and dividing by the volume of the culture in the reactor (300 mLs in our experiments). Figure 3.7 is presented here to highlight the

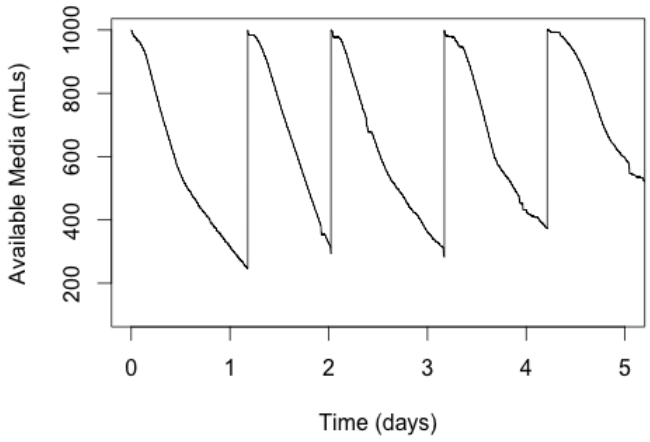


Figure 3.7. Available media during a directed evolution experiment of *Azotobacter vinelandii* in the turbidostat. Fresh media gets depleted as it is used to dilute the culture growing inside the reactor. A new media bottle is connected to the system as needed so that the reactor doesn't run out of media, which is seen here every time the available media returns to 1000 mLs (the use of 5 bottles of fresh media is shown). A representative number of days were selected to provide an example of the data the turbidostat collects while running.

functionality of the turbidostat and confirm that it was operating as anticipated to run directed evolution experiments going forward. Multiple turbidostats were attached to the computer concurrently and ran independently of one another.

During the proof of concept experiment where the *A. vinelandii* strain AZBB108 was grown in the presence of the toxic compound hydrazine, we were successful in obtaining a strain that could grow in the presence of elevated levels of hydrazine on plates (1 mM hydrazine, Figure 3.8) and 3 mM hydrazine in liquid. Although we were able to evolve AZBB108 to grow in the presence of

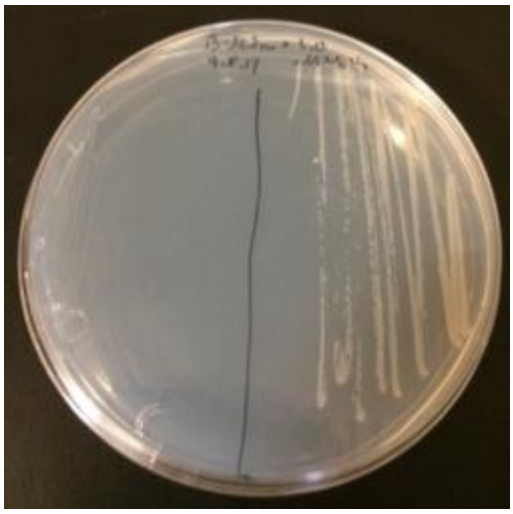


Figure 3.8. AZBB108 and evolved strain streaked onto Burke's (B) media plate with 1 mM hydrazine. AZBB108, the initial strain in the evolution experiment is streaked on the left side of the plate where there is no growth. The evolved strain that can grow in the presence of hydrazine is streaked on the right side of the plate. Image taken after 4 days of growth at 30º C.

hydrazine, we have not yet been successful in obtaining a strain with the ability to use hydrazine as its sole nitrogen source. When the atmosphere inside the reactor was switched from standard atmospheric conditions to an artificial atmosphere devoid of nitrogen (90% argon/10% O_2 by volume) but with hydrazine present, growth ceased. The culture was allowed time to evolve but eventually became contaminated. Although the ultimate goal of the specific directed evolution experiment has not been achieved to date, the overall goal showing that the turbidostat functioned as designed was confirmed.

FUTURE DIRECTIONS

Improvements can be made to our turbidostat system that could both lower building costs and make running experiments more user-friendly. Only recently has a laboratory published a paper describing a turbidostat cheaper than the one presented here ($350, Hoffman *et al.*, 2017). Less expensive stepper motor controllers are available and would be the easiest way to shave off about $100 from the final reactor price. A digital potentiometer could be used instead of the current utilized dial potentiometer to achieve more accurate media pump speeds. The software as designed is very simple, but is not the most intrinsically easy to use and would greatly benefit from having a user interface. Adding the ability to have the data automatically analyzed and emailed to the user instead of having to pull data from the computer and analyze it using separate software would also be ideal. Allowing for control and real-time feedback from the turbidostat through a mobile device app could also allow an experiment to run smoother. The biggest change under consideration to our current turbidostat design is to make a smaller version where the volume of the reactor is reduced from a 500 mL Pyrex bottle to a 10 mL test tube. Decreasing the size of the reactor would lower the overall cost per reactor and reduce bulk in the lab. It would also make it easier to run more replicates of an experiment or run different experiments in parallel. Both reactor setups would still be utilized as the original, larger turbidostat described here has the ability to contain more cells. A large number of cells is desirable in the context of an evolutionary experiment because subjecting more cells to a selective pressure increases the chances that a desired mutation in the population will occur.

Bibliography

Ainsworth EA, Bush DR. (2011) Carbohydrate export from the leaf: a highly regulated process and target to enhance photosynthesis and productivity. *Plant Physiol.* **155:**64-69.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389-402.

Avrahami-Moyal L, Engelberg D, Wenger JW, Sherlock G, Braun S. (2012) Turbidostat culture of *Saccharomyces cerevisiae* W303-1A under selective pressure elicited by ethanol selects for mutations in *SSD1* and *UTH1*. *FEMS Yeast Res*. **12**:521-533.

Barney BM, Eberhart LJ, Ohlert JM, Knutson CM, Plunkett MH. (2015) Gene Deletions Resulting in Increased Nitrogen Release by *Azotobacter vinelandii*: Application of a Novel Nitrogen Biosensor. *Appl Environ Microbiol.* **81**:4316-28.

Barney BM, Igarashi RY, Dos Santos PC, Dean DR, Seefeldt LC. (2004) Substrate Interaction at an Iron-Sulfur Face of the FeMo-cofactor during Nitrogenase Catalysis. *J Biol Chem*. **279**:53621-4.

Barney BM, Lee HI, Dos Santos PC, Hoffman BM, Dean DR, Seefeldt LC. (2006) Breaking the N_2 triple bond: insights into the nitrogenase mechanism. *Dalton Trans*. 2277-84.

Bihimidine S, Julius BT, Dweikat I, Braun DM. (2016) Tonoplast Sugar Transporters (SbTSTs) putatively control sucrose accumulation in sweet sorghum stems. *Plant Signal Behav.* **11**:e1117721.

Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J, Salamov A, Terry A, Yamada T, Dunigan DD, Grigoriev IV, Claverie JM, Van Etten JL. (2010) The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell*. **22**:2943-55.

Blumwald E, Tel-Or E. (1984) Salt Adaptation of the Cyanobacterium *Synechococcus* 6311 Growing in a Continuous Culture (Turbidostat). *Plant Physiol*. **74**:183-5.

Bosch TC. (2012) What hydra has to say about the role and origin of symbiotic interactions. *Biol Bull.* **223**:78-84

Brauer MJ, Huttenhower, C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell*. **19**:352-367.

Brechignac F, Schiller P. (1992) Pilot CELSS Based on a Maltose-Excreting *Chlorella*: Concept and Overview on the Technological Developments. *Adv Space Res*. **12**:33-6.

Burgess BK, Lowe DJ. (1996) Mechanism of Molybdenum Nitrogenase. *Chem Rev*. **96**:2983-3012

Chen H, Boutros PC. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. **12**:35.

Chen LQ, Cheung LS, Feng L, Tanner W, Frommer WB. (2015) Transport of Sugars. *Annu Rev Biochem*. **84:**865-94.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*. **10**:563-9.

Conant GC, Wolfe KH. (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*. **24**:861-2.

Concas A, Malavasi V, Costelli C, Fadda P, Pisu M, Cao G. (2016) Autotrophic growth and lipid production of *Chlorella sorokiniana* in lab batch and BIOCOIL photobioreactors: Experiments and modeling. *Bioresource Technol.* **211**:327-38.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**:3674-6.

Davis LC. (1980) Hydrazine as a substrate and inhibitor of *Azotobacter vinelandii* nitrogenase. *Arch Biochem Biophys.* **204**:270-6.

de-Bashan LE, Bashan Y, Moreno M, Lebsky VK, Bustillos JJ. (2002) Increased pigment and lipid content, lipid variety, and cell and population size of the microalgae *Chlorella* spp. when co-

immobilized in alginate beads with the microalgae-growth-promoting bacterium *Azospirillum brasilense*. *Can J Microbiol*. **48**:514-21.

Domozych DS, Ciancia M, Fangel JU, Mikkelsen MD, Ulvskov P, Willats WGT. (2012) The cell walls of green algae: a journey through evolution and diversity. *Front Plant Sci*. **3**:82.

Dorling M, McAuley PJ, Hodge H. (1997) Effect of pH on growth and carbon metabolism of maltose-releasing *Chlorella* (Chlorophyta). *Eur J Phycol*. **32**:19-24.

Dos Santos PC. (2011) Molecular Biology and Genetic Engineering in Nitrogen Fixation. *Methods Mol Biol*. **766**:81-92.

Douglas A, Smith DC. (1984) The Green Hydra Symbiosis VIII: Mechanisms in Symbiont Regulation. *Proc R Soc Ser B-Bio*. **221**:291-319.

Ducat DC, Avelar-Rivas JA, Way JC, Silver PA. (2012) Rerouting carbon flux to enhance photosynthetic productivity. *Appl Environ Microbiol*. **78**:2660-8.

Eom JS, Cho JI, Reinders A, Lee SW, Yoo Y, Tuan PQ, Choie SB, Bang G, Park YI, Cho MH, Bhoo SH, An G, Hahn RB, Ward JM, Jeon JS. (2011) Impaired function of the tonoplast-localized sucrose transporter in rice, OsSUT2, limits the transport of vacuolar reserve sucrose and affects plant growth. *Plant Physiol*. **157**:109-19.

Fischer A, Meindl D, Loos E. (1989) Glucose excretion by the symbiotic *Chlorella* of *Spongilla fluvaiatilis*. *Planta*. **179**:251-6.

Hengge-Aronis R. (1993) Survival of hunger and stress: The role of *rpoS* in early stationary phase gene regulation in *E. coli*. *Cell*. **72**:165-8.

Ho SH, Huang SW, Chen CY, Hasunuma T, Kondo A, Chang JS. (2013) Characterization and optimization of carbohydrate production from an indigenous microalga *Chlorella vulgaris* FSP-E. *Bioresource Technol*. **135**:157.65.

Hoffman BM, Lukoyanov D, Yang ZY, Dean DR, Seefeldt LC. (2014) Mechanism of nitrogen fixation by nitrogenase: the next stage. *Chem Rev*. **114**:4041-62.

Hoffman SA, Wohltat C, Muller KM, Arndt KM. (2017) A user-friendly, low-cost turbidostat with versatile growth rate estimation based on an extended Kalman filter. *PLoS One*. **12**:e0181923.

Hoshina R, Iwataki M, Imamura N. (2010) *Chlorella variabilis* and *Micractinium reisseri* sp nov (Chlorellaceae, Trebouxiophyceae): Redescription of the endosymbiotic green algae of *Paramecium bursaria* (Peniculia, Oligohymenophorea) in the 120th year. *Phycol Res*. **58**:188-201.

Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A. (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J*. **54**:621.39.

Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H, Horiuchi T. (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol.* **2**:2006-7.

Jung B, Ludewig F, Schulz A, Meibner G, Wostefeld N, Flugge, UI, Pommerrenig B, Wirsching P, Sauer N, Koch W, Sommer F, Muhlhaus T, Schroda M, Cuin TA, Graus D, Marten I, Hedrich R, Neuhaus HE. (2015) Identification of the transporter responsible for sucrose accumulation in sugar beet taproots. *Nat Plants*. **1:**14001.

Kessler E, Kauer G, Rahat M. (1991) Excretion of Sugars by *Chlorella* Species Capable and Incapable of Symbiosis with *Hydra viridis*. *Bot Acta*. **104**:58-63.

Kim D, Langmead B, Salzberg SL. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*. **12**:357-60.

Klumpp S, Zhang Z, Hwa T. (2009) Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell*. **139**:1366-1375.

Li L. Stoeckert CJ, Roos DS. (2003) OrthoMCL: Identification of orthologs groups for eukaryotic genomes. *Genome Res*. **13**:2178-9.

Li L. Zhang GQ, Wang QH. (2016) *De novo* transcriptomic analysis of *Chlorella sorokiniana* reveals differential genes expression in photosynthetic carbon fixation and lipid production. *BMC Microbiol*. **16**:223.

Lippincott J, Traxler B. (1997) MalFGK complex assembly and the transport and regulatory characteristics of MalK insertion mutants. *J Bacteriol*. **179**:1337-43.

Lu Y, Sharkey TD. (2006) The importance of maltose in transitory starch breakdown. *Plant Cell Environ*. **29**:353-66.

Manck-Gotzenberger J, Requena N. (2016) *Arbuscular mycorrhiza* Symbiosis Induces a Major Transcriptional Reprogramming of the Potato SWEET Sugar Transporter Family. *Front Plant Sci*. **7**:487.

Matteau D, Baby V, Pelletier S, Rodrigue S. (2015) A Small-Volume, Low-Cost, and Versatile Continuous Culture Device. *PLoS One.* **10**:e0133384.

Matzke B, Schwarzmeier E, Loos E. (1990) Maltose Excretion by the Symbiotic *Chlorella* of the Heliozoan *Acanthocystis turfacea. Planta.* **181**:593-8.

Mews LK, Smith DC. (1982) The Green Hydra Symbiosis VI: What Is the Role of Maltose Transfer from Alga to Animal? *Proc R Soc Ser B-Bio.* **216**:397-413

Oldham ML, Chen J. (2011) Crystal Structure of the Maltose Transporter in a Pretranslocation intermediate State. *Science.* **332**:1202-5.

Orsini M, Costelli C, Malavasi V, Cusano R, Concas A, Anglus A, Cao G. (2016) Complete sequence and characterization of mitochondrial and chloroplast genome of *Chlorella variabilis* NC64A. *Mitochondr DNA.* **27**:3128-30.

Price AL, Jones NC, Pevzner PA. (2005) De novo identification of repeat families in large genomes. *Bioinformatics*. **21**:351-8.

Pröschold T, Darienko T, Silva PC, Reisser W, Krienitz L. (2011) The systematics of *Zoochlorella* revisited employing an integrative approach. *Environ Microbiol.* **13**:350-64.

Sauer N. (2007) Molecular physiology of higher plant sucrose transporters. *FEBS Lett.* **581**:2309-17.

Schneider E, Hunke S. (1998) ATP-binding-cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. *Fems Microbiol Rev.* **22**:1-20.

Scofield GN, Hirose T, Aoki N, Furbank RT. (2007) Involvement of the sucrose transporter, OsSUT1, in the long-distance pathway for assimilate transport in rice. *J Exp Bot.* **58**:3155-69.

Setubal JC, dos Santos P, Goldman BS, Ertesvag H, Espin G, Rubio LM, Valla S, Almeida NF, Balasubramanian D, Cromes L, Curatti L, Du Z, Godsy E, Goodner B, Hellner-Burris K, Hernandez JA, Houmiel K, Imperial J, Kennedy C, Larson TJ, Latreille P, Ligon LS, Lu J, Maerk M, Miller NM, Norton S, O'Carrol IP, Paulsen I, Raulfs EC, Roemer R, Rosser J, Segura D, Slater S, Stricklin SL, Studholme DJ, Sun J, Viana CJ, Wallin E, Wang B, Wheller C, Zhu H, Dean DR, Dixon R, Wood D. (2009) Genome Sequence of *Azotobacter vinelandii*, an Obligate Aerobe Specialized To Support Diverse Anaerobic Metabolic Processes. *J Bacteriol* **191**:4534-45.

Shibata A, Takahashi F, Kasahara M, Imamura N. (2016) Induction of Maltose Release by Light in the Endosymbiont *Chlorella variabilis* of *Paramecium bursaria. Protist.* **167**:468-78.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* **31**:3210-2.

Stanke M, Diekhans M, Baertsch R, Haussler D. (2008) Using native and synthetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*. **24**:637-44.

Strandberg & Wilson. (1968) Formation of the nitrogen-fixing enzyme system in *Azotobacter vinelandii. Can J Microbiol.* **14**:25-31.

Summerer M, Sonntag B, Sommaruga R. (2007) An experimental test of the symbiosis specificity between the ciliate *Paramecium bursaria* and strains of the unicellular green alga *Chlorella. Environ Microbiol.* **9**:2117-22.

Takahashi CN, Miller AW, Ekness F, Dunham MJ, Klavins E. (2015) A low cost, customizable turbidostat for use in synthetic circuit characterization. *ACS Synth Biol.* **4**:32-8.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**:511-5.

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufo S, Li W. (2013) Prokaryotic Genome Annotation Pipeline. *NCBI Handbook*. **2**:173-86.

Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R. (2011) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat Genet*. **44**:101-5.

Villa JA, Ray EE, Barney BM. *Azotobacter vinelandii* siderophore can provide nitrogen to support the culture of the green algae *Neochloris oleoabundans* and *Scenedesmus* sp. BA032. *FEMS Microbiol Lett.* **351**:70-7.

Vinuesa P, Contreras-Moriera B. (2015) Robust Identification of Orthologues and Paralogues for Microbial Pan-Genomics Using GET_HOMOLOGUES: A Case Study of plncA/C Plasmids. *Methods Mol Biol.* **1231**:203-32.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One.* **9**:112963.

Wang Y, Chiu SY, Ho SH, Liu Z, Hasunuma T, Chang TT, Chang KF, Chang JS, Ren NQ, Kondo A. (2016) Improving carbohydrate production of *Chlorella sorokiniana* NIES-2168 through semi-continuous process coupled with mixotrophic cultivation. *Biotechnol J.* **11**:1072-81.

Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L. (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **31**:28-33.